Next VVSG
Part 1, Chapter 3:
Usability, Accessibility, and Privacy

December 12, 2007
Austin, TX

Dr. Sharon Laskowski
National Institute of Standards and Technology
sharon.laskowski@nist.gov

This presentation includes

- Key concepts for HFP requirements
- Overview of Part 1, Chapter 3
  - And relevant requirements in Part 1, 4.2.4, 4.4.2
- Explanations of the more complex requirements
  - Design requirements
  - Usability performance requirements
- Definitions as needed

Acronyms in this presentation

- Acc-VS: Accessible Voting Station
- DRE: Direct Recording Electronic
- EBM: Electronically-assisted Ballot Marker
- MMPB: Manually-Marked Paper Ballot
- VEBD: Voter-Editable Ballot Device
  - VEBD-A: Audio VEBD
  - VEBD-V: Video VEBD
- PCOS: Precinct-Count Optical Scanner
- ATI: Audio-Tactile Interface
- CVR: Cast Vote Record
- IVVR: Independent Voter-Verifiable Record
  - (VVPR: Voter-Verifiable Paper Record)
- VVPAT: Voter-Verifiable Paper Audit Trail
- CIF: Common Industry Format for Usability Test Reports
- VPP: Voting Performance Protocol

Key Concept: Goal

- Goal of these requirements is to provide a voting system that voters can use comfortably, efficiently, and with justified confidence that they have cast their votes correctly.
- Focus is primarily on the voter's interaction with the voting system, but there are requirements for poll workers.

Key Concept: Universal Usability

- View the voting station as a public kiosk, designed for “everyone”, not special purpose
- Many people have some special needs, but do not identify themselves as having disabilities
  - Aging population, those with language or reading issues
- Move as much accessibility into general voting station as possible
- **ALL usability requirements apply to the ACC-VS**

Key Concept: Accessibility

- Goal is to make the voting system independently accessible to as many voters as possible
- Harmonized with other accessibility standards, collaborated with the US Access Board
- Section 3.3 is organized according to the type of disability being addressed
  - NOTE: features intended primarily to address one kind of disability may assist voters with other kinds

Key Concept: Accessible Voter Verification

- Software independence (SI) for security and its implications for accessibility has been a thorny issue for the TGDC
  - SI in current systems can only be addressed using paper IVVR
  - But, paper by itself is not accessible: some voters cannot verify directly
  - So, there are a number of requirements that address how to make the IVVR accessible
  - This includes using “observational testing” to verify the reliability of indirect means of verification

Key Concept: Design vs. Performance

- Design requirements
  - Specify the “look and feel” of the voter interface, for general classes of voting systems
  - Based on best practice from other similar domains
  - Typically tested by inspection or expert review
- Performance requirements
  - Based on usability testing best practices
  - Specifies a benchmark which must be met when voters *interact* with the system
  - Tested in a tightly controlled environment with human test participants
  - We will explain the performance metrics and benchmarks in 3.2.1 in the final part of this presentation.

3.2 General Usability Requirements

To ensure manufacturer has a user-centered design and testing process, we have this and similar CIF reporting requirements:

3.2.1.2-A Usability testing by manufacturer for general population
"The manufacturer ***shall*** conduct summative usability tests on the voting system using individuals who are representative of the general population and ***shall*** report the test results, using the Common Industry Format, as part of the TDP."

3.2.2 Functional capabilities

This applies to all systems (VEBD and non-VEBD) and is modeled closely on HAVA provisions. For VEBD, voters can easily change their votes. For non-VEBD voters must obtain a new ballot to make changes.

3.2.2-A Notification of effect of overvoting
3.2.2-B Undervoting to be permitted
3.2.2-C Correction of ballot
3.2.2-D Notification of ballot casting

3.2.2.1 Editable interfaces

Voting systems such as DREs and EBMs present voters with an editable interface, allowing them to easily change their votes prior to final casting of the ballot.

3.2.2.1-A Prevention of overvotes
3.2.2.1-B Warning of undervotes
3.2.2.1-C Independent correction of ballot
3.2.2.1-D Ballot editing per contest
3.2.2.1-E Contest navigation
3.2.2.1-F Notification of ballot casting failure (DRE)

3.2.2.2 Non-Editable interfaces

Non-Editable interfaces, such as MMPB do not have the same flexibility as do editable interfaces. The PCOS gives feedback.

3.2.2.2-A Notification of overvoting
3.2.2.2-B Notification of undervoting
3.2.2.2-C Notification of blank ballots
3.2.2.2-D Ballot correction or submission following notification
3.2.2.2-E Handling of marginal marks
3.2.2.2-F Notification of ballot casting failure (PCOS)

3.2.3 Privacy

The voting process must preclude anyone else from determining the content of a voter's ballot without the voter's cooperation. Privacy ensures that the voter can cast votes based solely on his or her own preferences without intimidation or inhibition.

3.2.3.1 Privacy at the polls

3.2.3.1-A System support of privacy

3.2.3.1-A.1 Visual privacy

3.2.3.1-A.2 Auditory privacy

3.2.3.1-A.3 Privacy of warnings

3.2.3.1-A.4 No receipts

3.2.3.2 No recording of alternative format usage When voters use non-typical ballot interfaces, such as large print or alternative languages, their anonymity may be vulnerable. To the extent possible, only the logical contents of their ballots should be recorded in the CVR, not the special formats in which they were rendered.

3.2.3.2-A No recording of alternative languages

3.2.3.2-B No Recording of accessibility features

3.2.4 Cognitive issues

These requirements are intended to minimize cognitive difficulties for voters.

3.2.4-A Completeness of instructions
3.2.4-B Availability of assistance from the system

3.2.4-C Plain Language (Cognitive issues)

Although part of general usability, the use of plain language is also expected to assist voters with cognitive disabilities. The requirements apply to instructions that are inherent to the voting system.

3.2.4-C.1 Clarity of warnings
3.2.4-C.2 Context before action
3.2.4-C.3 Simple vocabulary
3.2.4-C.4 Start each instruction on a new line
3.2.4-C.5 Use of positive
3.2.4-C.6 Use of imperative voice
3.2.4-C.7 Gender-based pronouns

3.2.4-D No bias among choices (Cognitive issues)
Certain differences in presentation are mandated by state law, such as the order in which candidates are listed. However, comparable characteristics such as font size or voice volume and speed must be the same for all choices.

3.2.4-E Ballot design (Under 3.2.4 Cognitive issues)
That is, capability to design an easy-to-use ballot
- 3.2.4-E.1 Contests split among pages or columns
- 3.2.4-E.2 Indicate maximum number of candidates
- 3.2.4-E.3 Consistent representation of candidate selection
- 3.2.4-E.4 Placement of instructions

3.2.4-F Conventional use of color (Cognitive issues)
Note: As part of the test method work, a color guide is being developed, which will also address color blindness and contrast.

3.2.4-G Icons and language (Cognitive issues)
While icons can be used for emphasis when communicating with the voter, they must not be the sole means by which information is conveyed, since there is no widely accepted "iconic" language and therefore not all voters may understand a given icon.

3.2.5 Perceptual issues

These requirements are designed to minimize perceptual difficulties for the voter. Some are designed to assist voters with poor reading vision.

3.2.5-A Screen flicker
3.2.5-B Resetting of adjustable aspects at end of session
3.2.5-C Ability to reset to default values
3.2.5-D Minimum font size
3.2.5-E Available font sizes (VEBD-V)
3.2.5-F Use of sans serif font

3.2.5 Perceptual issues

3.2.5-G Legibility of paper ballots and verification records

While this requirement may be satisfied by one of its sub-requirements, other innovative solutions are not precluded.

3.2.5-G.1 Legibility via font size: "The system ***may*** achieve legibility of paper records by supporting the printing of those records in at least two font sizes, 3.0 - 4.0mm and 6.3 - 9.0mm."

3.2.5-G.2 Legibility via magnification: "The system ***may*** achieve legibility of paper records by supporting magnification of those records. This magnification ***may*** be done by optical or electronic devices. The manufacturer ***may*** either: 1) provide the magnifier itself as part of the system, or 2) provide the make and model number of readily available magnifiers that are compatible with the system."

Note: The magnifier(s) either provided or cited must, of course, provide legibility for the paper as actually presented on the system. For instance, if the paper record is under a transparent cover to prevent the voter from touching it, the means of magnification must be compatible with this configuration.

3.2.5 Perceptual issues

3.2.5-H Contrast Ratio
3.2.5-I High contrast for electronic displays
3.2.5-J Accommodation for color blindness
3.2.5-K No reliance solely on color

3.2.6 Interaction issues

Designed to minimize interaction difficulties for the voter.

- 3.2.6-A No page scrolling
- 3.2.6-B Unambiguous feedback for voter's selection
- 3.2.6-C Accidental Activation
  - 3.2.6-C.1 Size and separation of touch areas
  - 3.2.6-C.2 No repeating keys

Timing Issues: Definitions

Initial system response time: the time taken from when the voter performs some detectible action (such as pressing a button) to when the voting system *begins* responding in some obvious way (such as an audible response or any change on the screen).

Completed system response time: the time taken from when the voter performs some detectible action to when the voting system completes its response and settles into a stable state (e.g., finishes "painting" the screen with a new page).

Voter inactivity time: the amount of time from when the system completes its response until there is detectible voter activity. In particular, note that audio prompts from the system may take several minutes and that this time does not count as voter inactivity.

Alert time: the amount of time the equipment will wait for detectible voter activity after issuing an alert before going into an inactive state requiring poll worker intervention.

3.2.6.1 Timing issues

3.2.6.1-A Maximum initial system response time ≤ .5 secs

3.2.6.1-B Maximum completed system response time for vote confirmation ≤ 1 sec, ≤ 5 secs, audio

3.2.6.1-C Maximum completed system response time for all operations ≤ 10 secs, VEBD-V

3.2.6.1-D System response indicator .5 secs, if not complete in 1 sec

3.2.6.1-E Voter inactivity time 2-5 mins

3.2.6.1-F Alert time 20-45 secs

3.2.7 Alternative languages

These requirements are for voting systems to be *certified* for the languages declared by the manufacturer. Election officials must ensure that the voting system they deploy is capable of handling the languages meeting the legal VRA threshold within their districts.

3.2.7-A General support for alternative languages
- 3.2.7-A.1 Voter control of language
- 3.2.7-A.2 Complete information in alternative language
- 3.2.7-A.3 Auditability of records for English readers
- 3.2.7-A.4 Usability testing by manufacturer for alternative languages

3.2.8 Usability for poll workers

These are general capabilities that all systems must support for setup, operation, and shutdown.

3.2.8-A Clarity of system messages for poll workers

3.2.8.1 Operation

3.2.8.1-A Ease of normal operation
3.2.8.1-B Usability testing by manufacturer for poll workers
3.2.8.1-C Documentation usability
- 3.2.8.1-C.1 Poll Workers as target audience
- 3.2.8.1-C.2 Usability at the polling place
- 3.2.8.1-C.3 Enabling verification of correct operation

3.2.8.2 Safety

3.2.8.2-A Safety certification
"Equipment associated with the voting system *shall* be certified in accordance with the requirements of UL 60950-1, Information Technology Equipment – Safety – Part 1 by a certification organization accredited by the Department of Labor, Occupational Safety and Health Administration's Nationally Recognized Testing Laboratory program."

3.3 Accessibility Requirements

3.3.1 General

These requirements are relevant to a wide variety of disabilities.

3.3.1-A Accessibility throughout the voting session
3.3.1-A.1 Documentation of accessibility procedures
3.3.1-B Complete information in alternative formats
3.3.1-C No dependence on personal assistive technology
3.3.1-D Secondary means of voter identification
3.3.1-E Accessibility of paper-based vote verification
3.3.1-E.1 Audio readback for paper-based vote verification

Ensure that all voters have a similar opportunity for vote verification:

3.3.1-E Accessibility of paper-based vote verification

"If the Acc-VS generates a paper record (or some other durable, human-readable record) for the purpose of allowing voters to verify their votes, then the system ***shall*** provide a means to ensure that the verification record is accessible to all voters with disabilities, as identified in Part 1:3.3 "Accessibility requirements"."

Allow voters with visual disabilities to verify, even if indirectly, the contents of the record. The verification depends on the integrity of the mechanism that reads the record to the voter. The audio must be generated via the paper record and therefore not depend on any electronic or other "internal" record of the ballot.

3.3.1-E.1 Audio readback for paper-based vote verification

If the Acc-VS generates a paper record (or some other durable, human-readable record) for the purpose of allowing voters to verify their votes, then the system ***shall*** provide a mechanism that can read that record and generate an audio representation of its contents.

4.2.4-A IVVR vote-capture device, observational testing: IVVR vote-capture devices that support assistive technology ***shall*** support observational testing.

IVVR vote-capture device: Vote-capture device that achieves software independence through independent voter-verifiable records.

Vote-capture device: Device that is used directly by a voter to vote a ballot.

Software independence: Quality of a voting system or voting device such that a previously undetected change or fault in software cannot cause an undetectable change or error in election outcome.

Observational test: Operational test conducted on voting devices during an election, by real voters, to establish confidence that the VVPR is produced correctly when assistive technology is used. Discussion: Devices subjected to observational testing are used for normal collection of votes; the votes so collected are included in the election tally.

What about VVPAT (Voter-Verifiable Paper Audit Trail)?

- VVPAT is one implementation of paper IVVR, also called voter-verifiable paper records (VVPR)
- Section 4.4.2 contains requirements specific to VVPAT.
- All relevant usability and accessibility requirements apply to VVPAT.

3.3.2 Low vision

Section 3.2.5 applies as well (contrast, color-blindness, etc.)

3.3.2-A Usability testing by manufacturer for voters with low vision
3.3.2-B Adjustable saturation for color displays
3.3.2-C Distinctive buttons and controls
3.3.2-D Synchronized audio and video

3.3.3 Blindness

3.3.3-A Usability testing by manufacturer for blind voters
3.3.3-B Audio-tactile interface
  3.3.3-B.1 Equivalent functionality of ATI
  3.3.3-B.2 ATI supports repetition
  3.3.3-B.3 ATI supports pause and resume
  3.3.3-B.4 ATI supports transition to next or previous contest
  3.3.3-B.5 ATI can skip referendum wording

3.3.3 Blindness

3.3.3-C Audio features and characteristics
- 3.3.3-C.1 Standard connector
- 3.3.3-C.2 T-Coil coupling
- 3.3.3-C.3 Sanitized headphone or handset
- 3.3.3-C.4 Initial volume
- 3.3.3-C.5 Range of volume
- 3.3.3-C.6 Range of frequency
- 3.3.3-C.7 Intelligible audio
- 3.3.3-C.8 Control of speed

3.3.3 Blindness

3.3.3-D Ballot activation

3.3.3-E Ballot submission and vote verification

Purpose is that if voters using this station normally perform paper-based verification, or if they feed their own optical scan ballots into a reader, blind voters must also be able to do so.

3.3.3-F Tactile discernability of controls

3.3.3-G Discernability of key status

3.3.4 Dexterity

These specify the features of the accessible voting station designed to assist voters who lack fine motor control or use of their hands.

3.3.4-A Usability testing by manufacturer for voters with dexterity disabilities
3.3.4-B Support for non-manual input
3.3.4-C Ballot submission and vote verification
3.3.4-D Manipulability of controls
3.3.4-E No dependence on direct bodily contact

3.3.5 Mobility

Based on the ADA Accessibility Guidelines for Buildings and Facilities (ADAAG)

3.3.5-A Clear floor space
3.3.5-B Allowance for assistant
3.3.5-C Visibility of displays and controls
3.3.5.1Controls within reach
3.3.5.1-A Forward approach, no obstruction
3.3.5.1-B Forward approach, with obstruction
3.3.5.1-C Parallel approach, no obstruction
3.3.5.1-D Parallel approach, with obstruction

3.3.6 Hearing

3.3.6-A Reference to audio requirements
3.3.6-B Visual redundancy for sound cues
3.3.6-C No electromagnetic interference with hearing devices

3.3.7 Cognition

3.3.7-A General support for cognitive disabilities

The accessible voting station *should* provide support to voters with cognitive disabilities.

See other relevant requirements:

- Synchronization of audio with the displayed screen information (3.3.2-D)
- General cognitive usability requirements (3.2.4)
- Plain language (3.2.4-C)
- Large font sizes and legibility of paper (3.2.5-E, 3.2.5-G)
- Ability to control various aspects of the audio presentation (3.3.3-B, 3.3.3-C) such as pausing, repetition, and speed.

3.3.8 English proficiency

3.3.8-A Use of ATI

For voters who lack proficiency in reading English, the voting equipment *shall* provide an audio interface for instructions and ballots as described in Part 1:3.3.3-B.

3.3.9 Speech

3.3.9-A Speech not to be required by equipment

Usability Performance Requirements

- Goal: To develop a test method to distinguish systems with poor usability from those with good usability
  - Based on performance not evaluation of the design
  - Reliably detects and counts errors one might see when voters **interact** with a voting system
  - Reproducible by test laboratories
  - Technology-independent

Calculating benchmarks

- Given such a test method, benchmarks can be calculated: a system meeting the benchmarks has good usability and passes the test
  - The values chosen for the benchmarks become the performance requirements

Usability testing for certification in a lab

- We are measuring the **performance of the system** in a **lab**
  - We control for other variables, including the test participants
  - We measure the effect of the system on usability
- The test ballot is designed to detect different types of usability errors and be typical of many types of ballots
- The test environment is tightly controlled, e.g., for lighting, setup, instructions, no assistance
- The test participants are chosen to reliably detect the same performance on the same system

Usability testing for certification in a lab

- Test participants are told exactly how to vote, so errors can be measured
- The test results measure relative degree of usability between systems and are NOT intended to predict performance in a specific election
  - Ballot is different
  - Environment is different (e.g, help is provided)
  - Voter demographics are different
- A general sample of the US voting population is never truly representative because all elections are “local”.

Components of the test method (Voting Performance Protocol)

- Well-defined test protocol that describes the number and characteristics of the "voters" participating in the test and how to conduct test,
- Test ballot that is relatively complex to ensure the entire voting system is evaluated and significant errors detected,
- Instructions to the "voters" on exactly how to vote so that errors can be accurately counted,
- Description of the test environment,
- Method of analyzing and reporting the results, and
- Performance benchmarks with associated threshold values.

Performance Benchmarks: Recap of Research

- Validity: tested on 2 different systems with 47 participants
  - Test protocol detected differences between systems, produces errors that were expected.
- Repeatability/Reliability: 4 tests on same system, 195 participants, similar results

Performance Benchmarks: Recap of Research

- Demographics
  - Eligible to vote in the US
  - Gender: 60% female , 40% male
  - Race: 20 % African American, 70% Caucasian, 10% Hispanic
  - Education: 20 % some college, 50% college graduate, 30% post graduate
  - Age: 30% 25-34 yrs., 35% 35-44 yrs., 35 % 45-54 yrs.
  - Geographic Distribution: 80% VA, 10% MD, 10% DC

Benchmark Tests

- 4 systems, May 19-20, June 1-2
  - Selection of DREs, EBMs, PCOS
- 187 test participants
- 5 measurements
  - 3 benchmark thresholds
  - 2 values to be reported only

The Performance Measures Base Accuracy Score

- We first count the number of errors test participants made on the test ballot – there are 28 voting opportunities: count how many were correct for each participant
- We then calculate a **Base Accuracy Score:** the mean percentage of all ballot choices that are correctly cast by the test participants

We calculate 3 effectiveness measures: Total Completion Score

- The percentage of test participants who were able to complete the process of voting and have their ballot choices recorded by the system.

Voter Inclusion Index (VII)*

- A measure of overall voting accuracy that uses the Base Accuracy Score and the standard deviation.
  - If 2 systems have the same Base Accuracy Score (BAS), the system with the larger variability gets a lower VII.
  - The formula, where S is the standard deviation and LSL is a lower specification limit to spread out the measurement (we used .85), is:

$$VII = \frac{BAS - LSL}{3S}$$

*range is 0 to ~1, assuming best value is 100% BAS, S=.05, but may be higher

Perfect Ballot Index (PBI)*

- The ratio of the number of cast ballots containing no erroneous votes to the number of cast ballots containing at least one error.
  - This measure deliberately magnifies the effect of even a single error. It identifies those systems that may have a high Base Accuracy Score, but still have at least one error made by many participants.
  - This might be caused by a single voting system design problem, causing a similar error by the participants. The higher the value of the index, the better the performance of the system.

*range is 0 to infinity, if no errors at all.

Efficiency and Confidence Measures

- **Average Voting Session Time** – mean time taken for test participants to complete the process of activating, filling out, and casting the ballot.
- **Average Voter Confidence** – mean confidence level expressed by the voters that they believed they voted correctly and the system successfully recorded their votes.
- Neither of these measures were correlated with effectiveness.
- Most people were confident in the system and their ability to use the system.

Benchmark test results

	Number of Participants Completing The Ballot	Total Completion Score (%) Confidence Intervals (95 % level)	Base Accuracy Score (%) Mean, Standard Deviation	Voter Inclusion Index With 85% LSL Confidence Intervals (95 % level)
System A	50 of 52 (96.2%)	86.3-99.7	95.0, 11	.19-.41
System B	42 of 42 (100%)	92.8-100	96.0, 6	.49-.85
System C	43 of 43 (100%)	92.9-100	92.4, 13	.08-.30
System D	47 of 50 (94.0%)	83.2-98.6	92.4, 19	.03-.22

Benchmark test results

	Number of Participants with Perfect Ballot Including Percent and Index using Adjusted Wald Method	Perfect Ballot Index Confidence Intervals (95 % level)	Voting Time (secs) Mean, Standard Deviation	Participant Confidence (1-5) Mean, Standard Deviation
System A	29 of 50 (58.0%) Index: 1.35	0.79 – 2.40	638.1, 166.1	4.0, 1.0
System B	24 of 42 (57.1%) Index: 1.30	0.73 – 2.44	429.3, 156.3	3.3, 1.4
System C	15 of 43 (34.9%) Index: 0.57	0.29 – 1.00	870.7, 236.0	3.6, 1.4
System D	31 of 47(66%) Index: 1.84	1.07 – 3.52	744.7, 209.3	3.8, 1.2

Benchmark thresholds

- Voting systems, when tested by laboratories designated by the EAC using the methodology specified in this paper, must meet or exceed ALL these benchmarks:
  - Total Completion Score of 98%
  - Voter Inclusion Index of .35
  - Perfect Ballot Index of 2.33
- Systems C and D fail.
- Report time and confidence

3.2.1.1-A Total completion performance: The system *shall* achieve a total completion score of at least 98% as measured by the VPP.

3.2.1.1-B Perfect ballot performance: The system *shall* achieve a perfect ballot index of at least 2.33 as measured by the VPP.

3.2.1.1-C Voter inclusion performance: The system *shall* achieve a voter inclusion index of at least 0.35 as measured by the VPP.

3.2.1.1-D Usability metrics from the Voting Performance Protocol: The test lab *shall* report the metrics for usability of the voting system, as measured by the VPP.

3.2.1.1-D.1 Effectiveness metrics for usability: The test lab *shall* report all the effectiveness metrics for usability as defined and measured by the VPP.

3.2.1.1-D.2 Voting session time: The test lab *shall* report the average voting session time, as measured by the VPP.

3.2.1.1-D.3 Average voter confidence: The test lab *shall* report the average voter confidence, as measured by the VPP.

How "tough" should the benchmark thresholds be?

- The benchmark data here used 50 test participants, but the test protocol will call for 100 (to allow statistical assumption of normal distribution to calculate the VII confidence intervals)
  - 100 participants will narrow the confidence intervals and thereby toughen the test.
- **Two points of view:**
  - Proposed benchmarks do weed out poorly performing systems (and, it is relatively easy to raise thresholds)

  vs.
  - This should be a forward-looking standard, new systems should be held to a higher standard
    - (but what is the upper bound, given that humans always make some mistakes?)

Additional Research

- Reproducibility: How much flexibility can be allowed in the test protocol?
  - Will variability in test participants experience due to labs in different geographic regions affect results?
  - Should we factor in older population or less educated population?
  - Benchmark thresholds are always tied to the demographics of the test participants to some extent
- Accessible voting system performance?

Any Final Questions?