Technical Guidelines Development Committee
March 22, 2007, Plenary Meeting

Review of CRT Changes

Technical Guidelines Development Committee
March 22, 2007, Plenary Meeting

Terminology Standard

- Mean Time Between Failure / failure rate
- Election official
- Electronically-assisted Ballot Marker
- Candidate and choice
- Reporting context

Technical Guidelines Development Committee
March 22, 2007, Plenary Meeting

Product Standard

- Principal criteria
- General requirements
- Benchmarks
- Coding conventions
- Archivalness of media
- Tabulation integrity
- Logic model

Standards on Data to be Provided

- User documentation
- Test report and Public Information Package

Testing Standard

- Documentation and design reviews
- Functional testing
- Benchmarks (test methods)

Technical Guidelines Development Committee
March 22, 2007, Plenary Meeting

Benchmarks

David Flater
Computer Scientist

What is a benchmark?

- Definition: Quantitative point of reference to which the measured performance of a system or device may be compared
- Plain language: The number specified in the requirement (e.g., the failure rate shall not exceed [*benchmark*])
- The VVSG contains benchmarks for:
  - Reliability (failure rate)
  - Accuracy (error rate)
  - Rate of misfeeds for paper-based tabulators

Technical Guidelines Development Committee
March 22, 2007, Plenary Meeting

Problems to fix in next VVSG

- Existing Mean Time Between Failure (MTBF) benchmark condemned in public comments and thrown out by resolution of TGDC
  - Need a new benchmark
- Existing accuracy benchmark is ambiguous (see discussion paper "On Accuracy Benchmarks, Metrics, and Test Methods" from December 2006 Plenary Meeting)
  - At a minimum, need confirmation that the drafted clarification is acceptable
  - Changing the numbers is an option
- Existing misfeed rate benchmark condemned in CRT discussion
  - Already changed based on CRT input

Technical Guidelines Development Committee
March 22, 2007, Plenary Meeting

From December meeting

- Input needed from election officials
  - Acceptable % failures: 0 % .. 30 %
  - Acceptable # errors: 0 .. 1000
  - Volumes for each type of device

Technical Guidelines Development Committee
March 22, 2007, Plenary Meeting

Status

- 2007-01-17: Sent letters to NASS and NASED
- 2007-02-08: NASS declined to take a position
- 2007-03-09: Received response from NASED (available at http://vote.nist.gov/ECPosStat.htm)
- 2007-03-15: Discussed on CRT teleconference
- 2007-03-16: Deadline for presentations for this meeting

Technical Guidelines Development Committee
March 22, 2007, Plenary Meeting

Paraphrasing NASED on Reliability

- No failures that lead to unrecoverable votes are acceptable
- In other cases, our tolerance for failures depends on how hard it is to recover from those failures
- There is no “typical” volume on which to base a benchmark
- Five categories of reliability / things that need to happen

Reliability by design	Vendor quality assurance, volume test
Resilience to and avoidance of human error	Usability testing, volume test
Manufacturing quality	Vendor quality control, volume test
Longevity	Shake-and-bake tests, full life-cycle monitoring
Maintainability	Unclear; “cost of ownership”

Consequences for Benchmark (1)

- To empower test labs to advise rejection of systems that perform unreliably during testing, there needs to be a benchmark for what constitutes an unacceptable rate of failure
- So even though the right answer in practice depends on so many things, and there is no “typical” volume, we still need a number
- Zero?
- (Full circle)

Consequences for Benchmark (2)

- We cannot know at certification time what the practical impact of different sorts of failures will be—it depends on practices and procedures (circular reasoning)
- We could assign different weights to different kinds of failures if we could define them in an objectively determinable way
- The 1990 VSS tried to do this (Appendix G, Voting System Failure Definition and Scoring Criteria)
- All that was removed in the 2002 VSS
- As of the deadline for this presentation, Paul Miller is following up to find out why

Paraphrasing NASED on Accuracy

- The acceptable number of errors is one less than the vote margin between first and second place
- On 1 in 10 000 000 (ballot positions) benchmark
  - Is a compromise based on cost of testing
  - "No reason to change," but…
  - Recognize need to review test methods
  - Achievable for perfect test ballots but maybe not for real ballots

Consequences for Benchmark

- By the vote margin criterion, zero errors isn't good enough
- Start with 1 in 10 000 000 (ballot positions)
- Neutral on ballot positions versus report total error rate, but this definition may impact the propriety of the number
- Since volume testing should produce "real" ballots, the benchmark should be relaxed to what is achievable using "real" ballots—unless you want to disqualify large categories of systems
- We do not have that figure
- (Full circle)

Technical Guidelines Development Committee
March 22, 2007, Plenary Meeting

The Bottom Line

- Diligence thus far has not yielded the data necessary to derive defensible benchmarks
- Time is short
- We need all input—TGDC, NASED, public comments to voting@nist.gov—ASAP
- Failing that, the text that is in the draft now will remain substantively unchanged in the next draft
- Public and EAC review process is the last chance to make changes