AUDITS: THE AFTER-MATH OF ELECTION REFORM

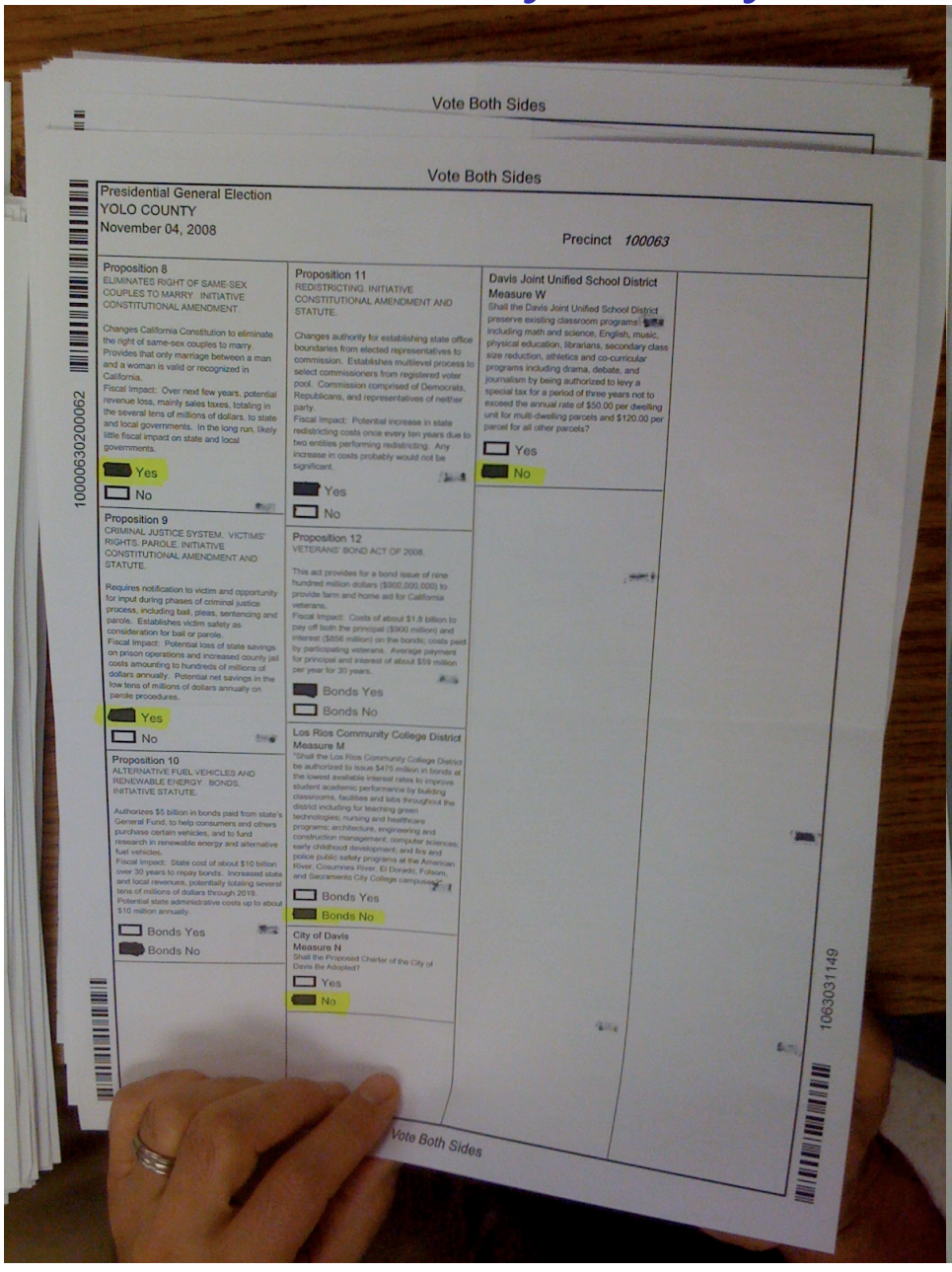
Conference on Innovative Electoral Reforms and Strategies

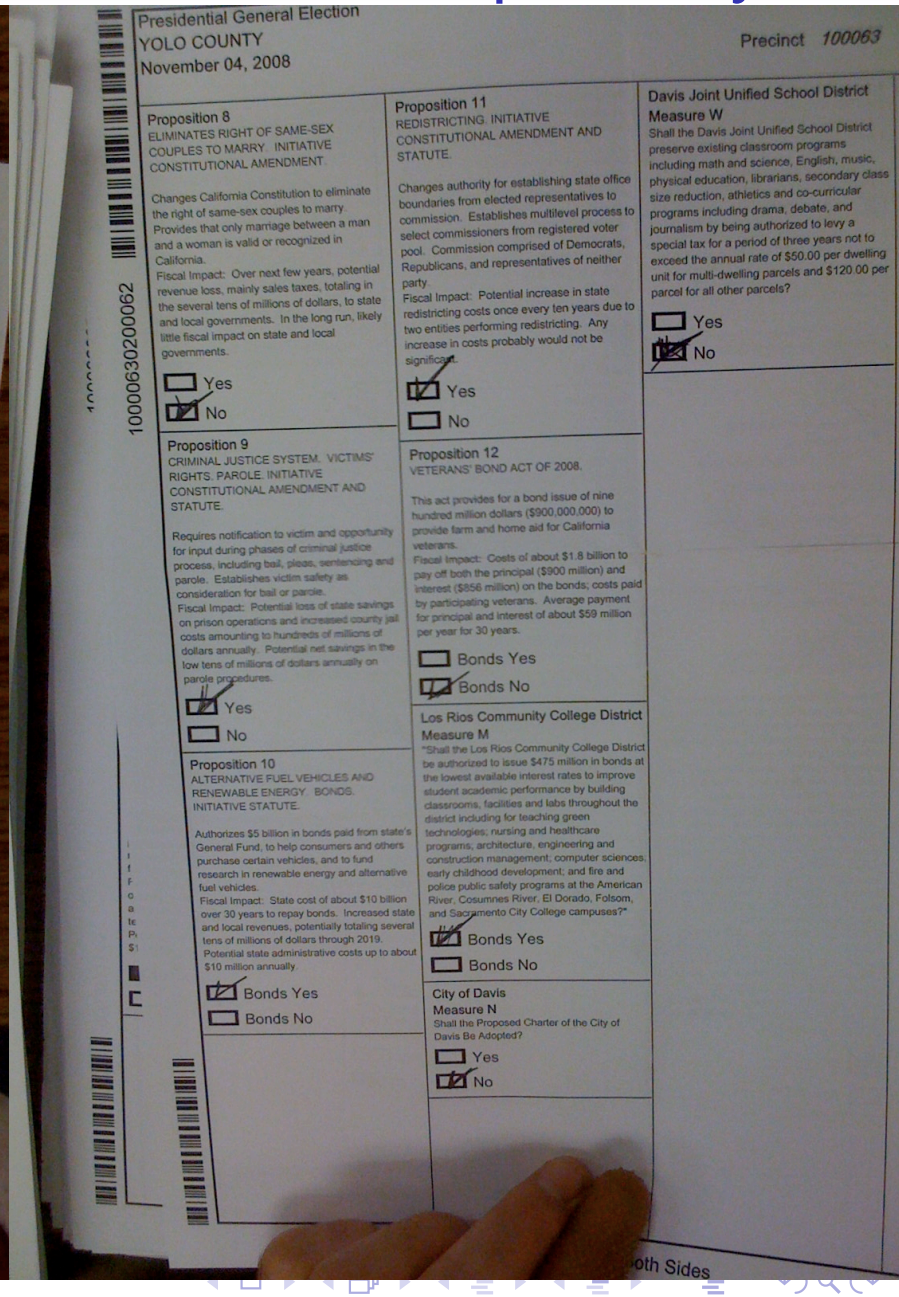
Philip B. Stark

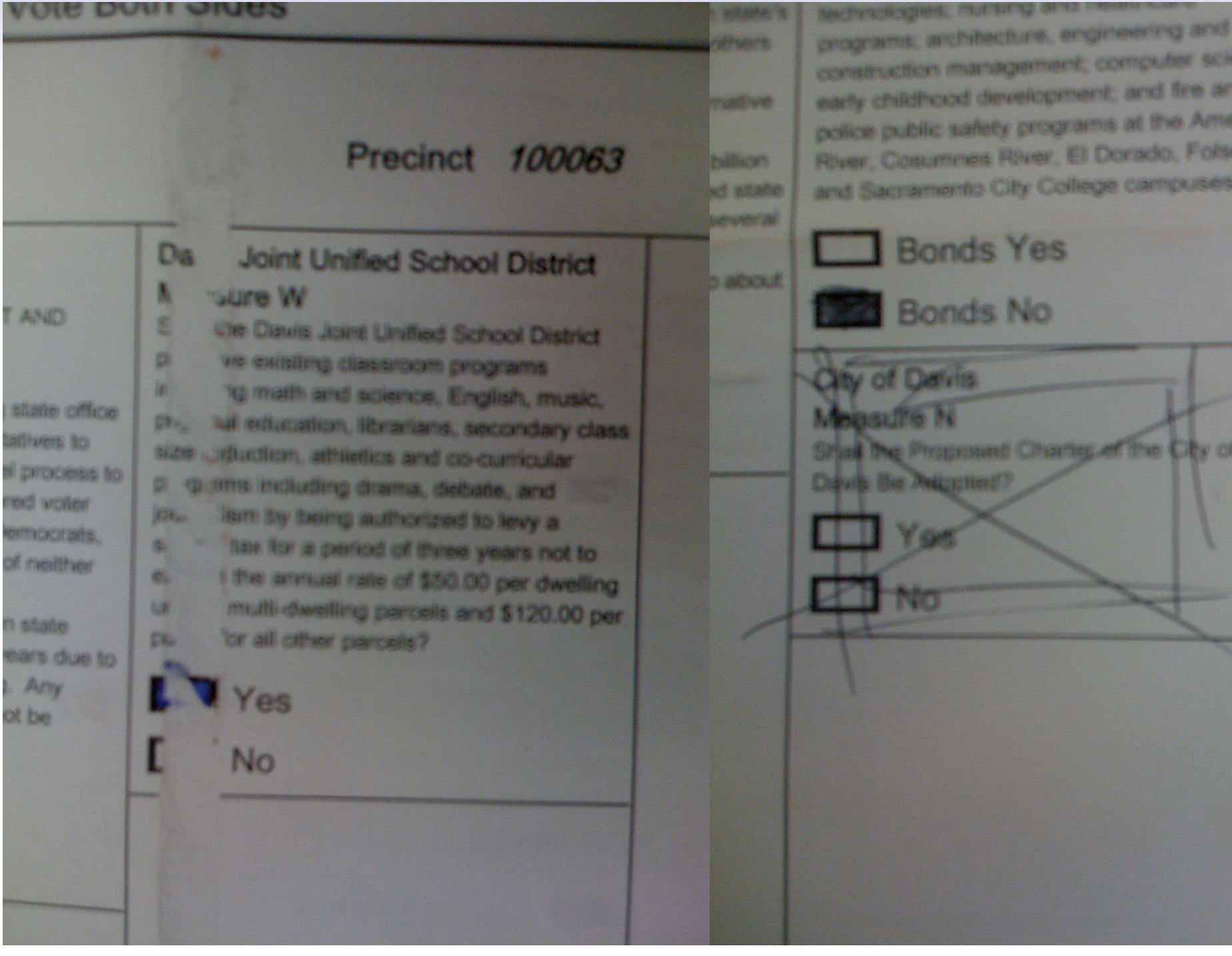
Department of Statistics
University of California, Berkeley

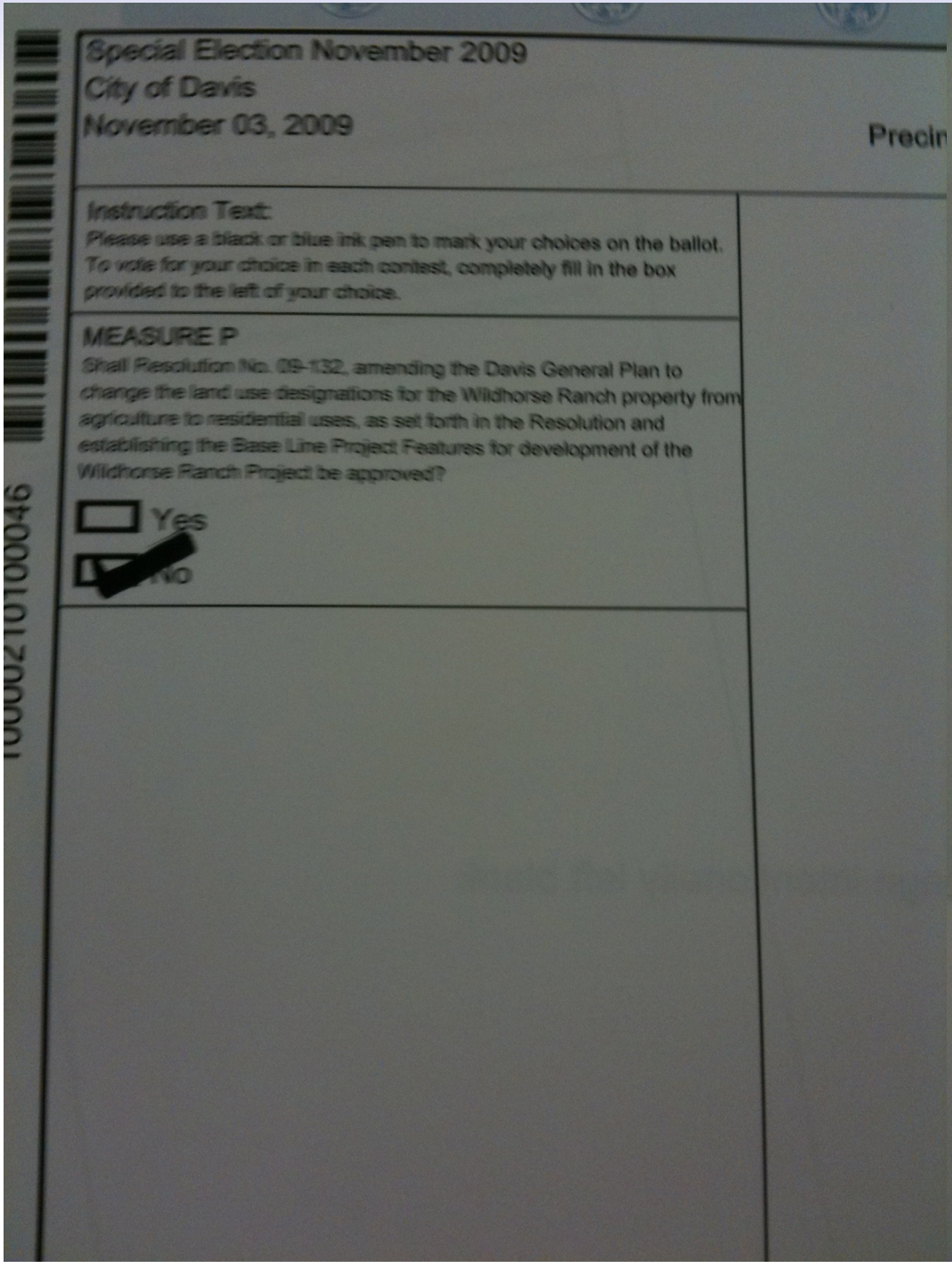
10–11 December 2010

Voters marks vary widely: No VTS can count perfectly






Vote Both Sides
Precinct 100063
Joint Unified School District
Yes
No
Bonds Yes
Bonds No
City of Davis
Measure N
Yes
No


Special Election November 2009
City of Davis
November 03, 2009

Precin

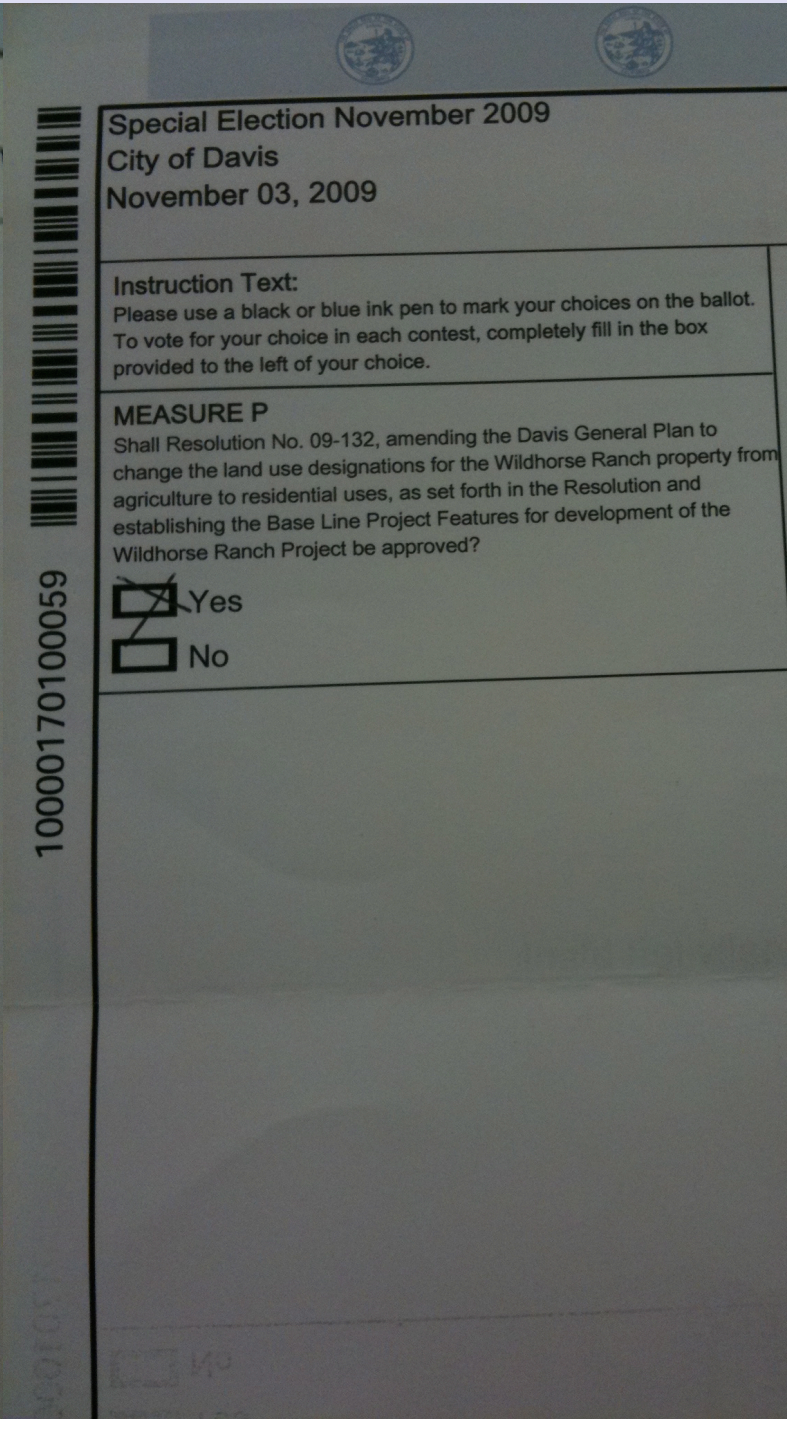
Instruction Text:
Please use a black or blue ink pen to mark your choices on the ballot. To vote for your choice in each contest, completely fill in the box provided to the left of your choice.

MEASURE P
Shall Resolution No. 09-132, amending the Davis General Plan to change the land use designations for the Wildhorse Ranch property from agriculture to residential uses, as set forth in the Resolution and establishing the Base Line Project Features for development of the Wildhorse Ranch Project be approved?

Yes

No

10000210100046


Special Election November 2009
City of Davis
November 03, 2009

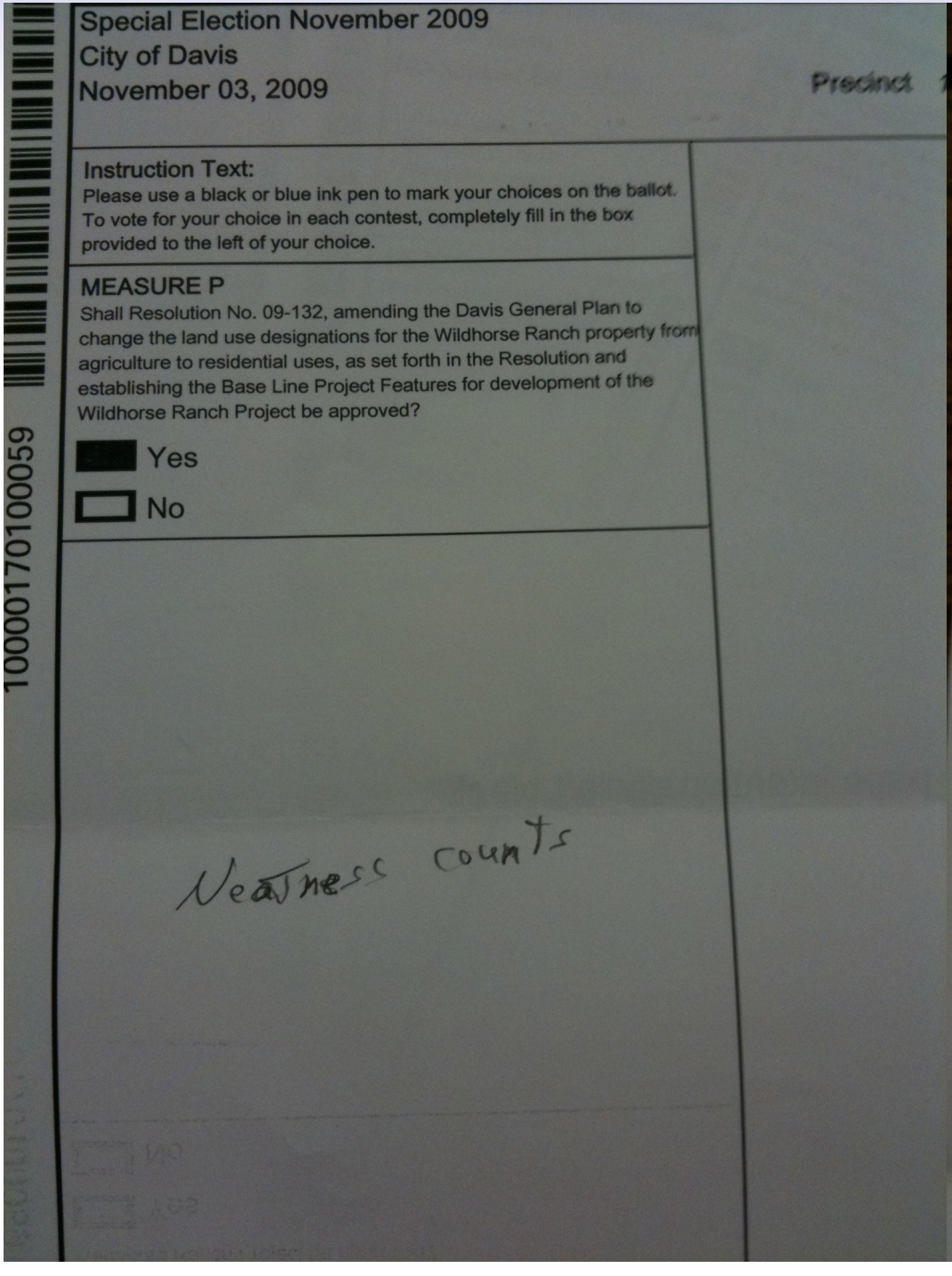
Instruction Text:
Please use a black or blue ink pen to mark your choices on the ballot. To vote for your choice in each contest, completely fill in the box provided to the left of your choice.

MEASURE P
Shall Resolution No. 09-132, amending the Davis General Plan to change the land use designations for the Wildhorse Ranch property from agriculture to residential uses, as set forth in the Resolution and establishing the Base Line Project Features for development of the Wildhorse Ranch Project be approved?

Yes

No

10000170100059



Special Election November 2009
City of Davis
November 03, 2009

Precinct

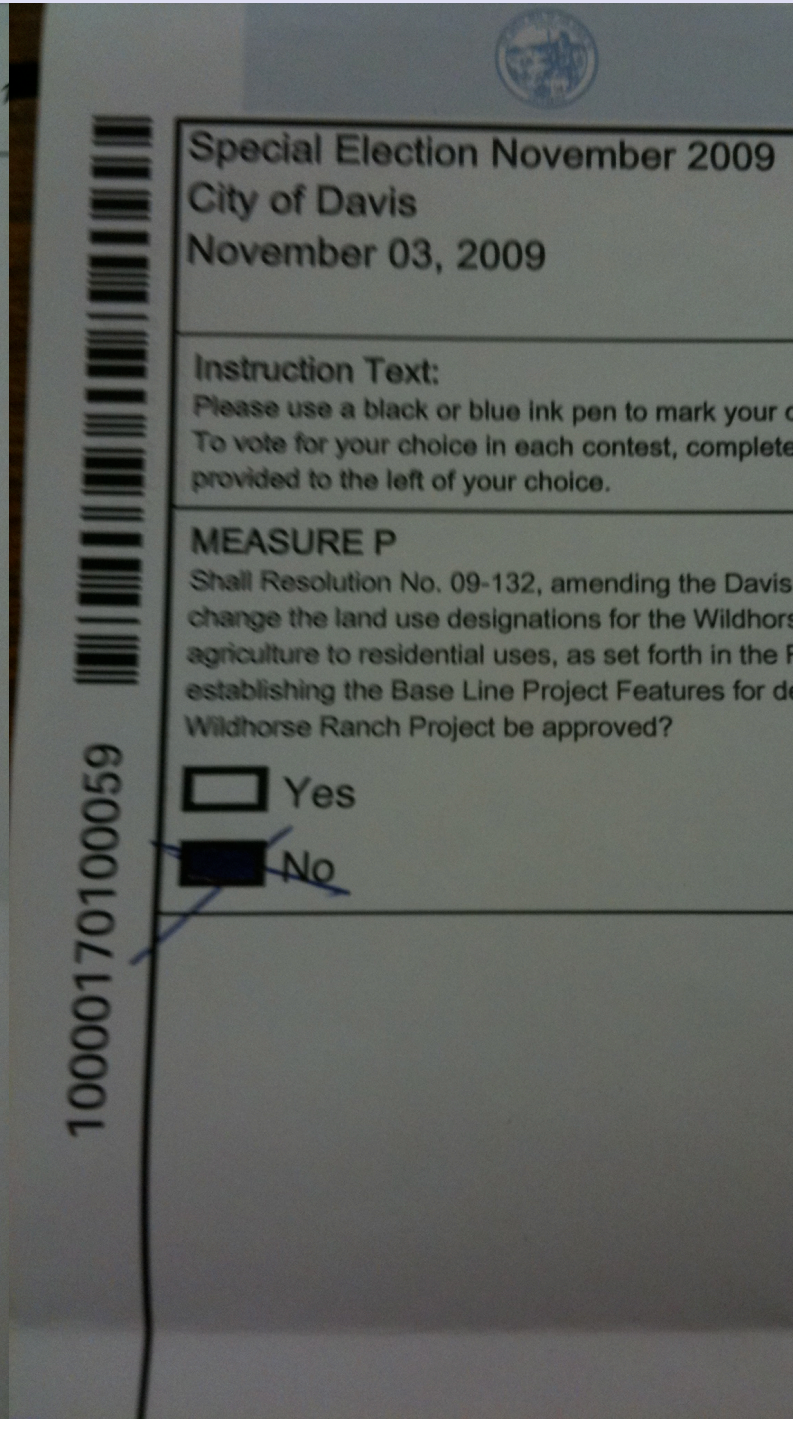
Instruction Text:
Please use a black or blue ink pen to mark your choices on the ballot. To vote for your choice in each contest, completely fill in the box provided to the left of your choice.

MEASURE P
Shall Resolution No. 09-132, amending the Davis General Plan to change the land use designations for the Wildhorse Ranch property from agriculture to residential uses, as set forth in the Resolution and establishing the Base Line Project Features for development of the Wildhorse Ranch Project be approved?

■ Yes
☐ No

Neatness counts

10000170100059



Special Election November 2009
City of Davis
November 03, 2009

Instruction Text:
Please use a black or blue ink pen to mark your c
To vote for your choice in each contest, complete
provided to the left of your choice.

MEASURE P
Shall Resolution No. 09-132, amending the Davis
change the land use designations for the Wildhors
agriculture to residential uses, as set forth in the
establishing the Base Line Project Features for d
Wildhorse Ranch Project be approved?

☐ Yes
■ No

10000170100059

Why audit?

- Any way of counting votes makes mistakes.
- If there are enough mistakes, apparent winner could be wrong.
- If there's a complete, accurate audit trail, can ensure big chance of correcting wrong outcomes—but you have to count by hand.
- Other reasons too: process improvement, deterrence, etc.
- Compliance audits vs. materiality audits.

Why audit?

- Any way of counting votes makes mistakes.
- If there are enough mistakes, apparent winner could be wrong.
- If there's a complete, accurate audit trail, can ensure big chance of correcting wrong outcomes—but you have to count by hand.
- Other reasons too: process improvement, deterrence, etc.
- Compliance audits vs. materiality audits.

Why audit?

- Any way of counting votes makes mistakes.
- If there are enough mistakes, apparent winner could be wrong.
- If there's a complete, accurate audit trail, can ensure big chance of correcting wrong outcomes—but you have to count by hand.
- Other reasons too: process improvement, deterrence, etc.
- Compliance audits vs. materiality audits.

Why audit?

- Any way of counting votes makes mistakes.
- If there are enough mistakes, apparent winner could be wrong.
- If there's a complete, accurate audit trail, can ensure big chance of correcting wrong outcomes—but you have to count by hand.
- Other reasons too: process improvement, deterrence, etc.
- Compliance audits vs. materiality audits.

Why audit?

- Any way of counting votes makes mistakes.
- If there are enough mistakes, apparent winner could be wrong.
- If there's a complete, accurate audit trail, can ensure big chance of correcting wrong outcomes—but you have to count by hand.
- Other reasons too: process improvement, deterrence, etc.
- Compliance audits vs. materiality audits.

Crucial ingredients

- Essential that voters create complete, durable, accurate audit trail.
- Essential that voting systems enable auditors to access reported results (total ballots, counts for each candidate, registered voters) in auditable batches. (Smaller batches are better.)
- Essential to select batches at random, *after* the results are posted. (Can supplement with "targeted" samples.)
- Need a plan for dealing with discrepancies, possibly leading to full count. "Reconciling" or "explaining" isn't enough.

Crucial ingredients

- Essential that voters create complete, durable, accurate audit trail.
- Essential that voting systems enable auditors to access reported results (total ballots, counts for each candidate, registered voters) in auditable batches. (Smaller batches are better.)
- Essential to select batches at random, *after* the results are posted. (Can supplement with "targeted" samples.)
- Need a plan for dealing with discrepancies, possibly leading to full count. "Reconciling" or "explaining" isn't enough.

Crucial ingredients

- Essential that voters create complete, durable, accurate audit trail.
- Essential that voting systems enable auditors to access reported results (total ballots, counts for each candidate, registered voters) in auditable batches. (Smaller batches are better.)
- Essential to select batches at random, *after* the results are posted. (Can supplement with "targeted" samples.)
- Need a plan for dealing with discrepancies, possibly leading to full count. "Reconciling" or "explaining" isn't enough.

Crucial ingredients

- Essential that voters create complete, durable, accurate audit trail.
- Essential that voting systems enable auditors to access reported results (total ballots, counts for each candidate, registered voters) in auditable batches. (Smaller batches are better.)
- Essential to select batches at random, *after* the results are posted. (Can supplement with "targeted" samples.)
- Need a plan for dealing with discrepancies, possibly leading to full count. "Reconciling" or "explaining" isn't enough.

New York's Recent Reforms

Moved to precinct-count optically scanned paper ballots.

Introduced audit laws, starting with 3% of machines (scanners).

Irreconcilable differences between hand count and machine count can lead to counting more ballots by hand: 5%, 12%, or all.

NY SD 7

- Balance of power in NY Senate: Either 31 seats for each party, or 32 for Republicans.
- Reported margin of 451 votes (0.5%) for Republican candidate Martins.
- Disagreement about purpose and requirements of auditing.

NY SD 7

- Balance of power in NY Senate: Either 31 seats for each party, or 32 for Republicans.
- Reported margin of 451 votes (0.5%) for Republican candidate Martins.
- Disagreement about purpose and requirements of auditing.

NY SD 7

- Balance of power in NY Senate: Either 31 seats for each party, or 32 for Republicans.
- Reported margin of 451 votes (0.5%) for Republican candidate Martins.
- Disagreement about purpose and requirements of auditing.

NY SD 7: Audit

- 7 of 249 of machines audited at random (3%).
- 3 of 7 (i.e., 43%) showed errors. Net error favored the apparent winner.
- Republicans: the errors were “reconciled”: Machines are fine. Democrats disagree.
- Judge Warshawsky: “In my opinion, reconcilable would be ‘Is there a clear reason why the deficiency occurred?’ ”
- Is “the machine was mis-programmed” a clear reason?
- In my opinion, reconcilable would be “We counted again by hand and found that the error was in the hand count, not the machine count.”
- “Clear reason” is irrelevant for whether the apparent outcome of the contest is correct. Size of the difference matters.

NY SD 7: Audit

- 7 of 249 of machines audited at random (3%).
- 3 of 7 (i.e., 43%) showed errors. Net error favored the apparent winner.
- Republicans: the errors were “reconciled”: Machines are fine. Democrats disagree.
- Judge Warshawsky: “In my opinion, reconcilable would be ‘Is there a clear reason why the deficiency occurred?’ ”
- Is “the machine was mis-programmed” a clear reason?
- In my opinion, reconcilable would be “We counted again by hand and found that the error was in the hand count, not the machine count.”
- “Clear reason” is irrelevant for whether the apparent outcome of the contest is correct. Size of the difference matters.

NY SD 7: Audit

- 7 of 249 of machines audited at random (3%).
- 3 of 7 (i.e., 43%) showed errors. Net error favored the apparent winner.
- Republicans: the errors were “reconciled”: Machines are fine. Democrats disagree.
- Judge Warshawsky: “In my opinion, reconcilable would be ‘Is there a clear reason why the deficiency occurred?’ ”
- Is “the machine was mis-programmed” a clear reason?
- In my opinion, reconcilable would be “We counted again by hand and found that the error was in the hand count, not the machine count.”
- “Clear reason” is irrelevant for whether the apparent outcome of the contest is correct. Size of the difference matters.

NY SD 7: Audit

- 7 of 249 of machines audited at random (3%).
- 3 of 7 (i.e., 43%) showed errors. Net error favored the apparent winner.
- Republicans: the errors were “reconciled”: Machines are fine. Democrats disagree.
- Judge Warshawsky: “In my opinion, reconcilable would be ‘Is there a clear reason why the deficiency occurred?’ ”
- Is “the machine was mis-programmed” a clear reason?
- In my opinion, reconcilable would be “We counted again by hand and found that the error was in the hand count, not the machine count.”
- “Clear reason” is irrelevant for whether the apparent outcome of the contest is correct. Size of the difference matters.

NY SD 7: Audit

- 7 of 249 of machines audited at random (3%).
- 3 of 7 (i.e., 43%) showed errors. Net error favored the apparent winner.
- Republicans: the errors were “reconciled”: Machines are fine. Democrats disagree.
- Judge Warshawsky: “In my opinion, reconcilable would be ‘Is there a clear reason why the deficiency occurred?’ ”
- Is “the machine was mis-programmed” a clear reason?
- In my opinion, reconcilable would be “We counted again by hand and found that the error was in the hand count, not the machine count.”
- “Clear reason” is irrelevant for whether the apparent outcome of the contest is correct. Size of the difference matters.

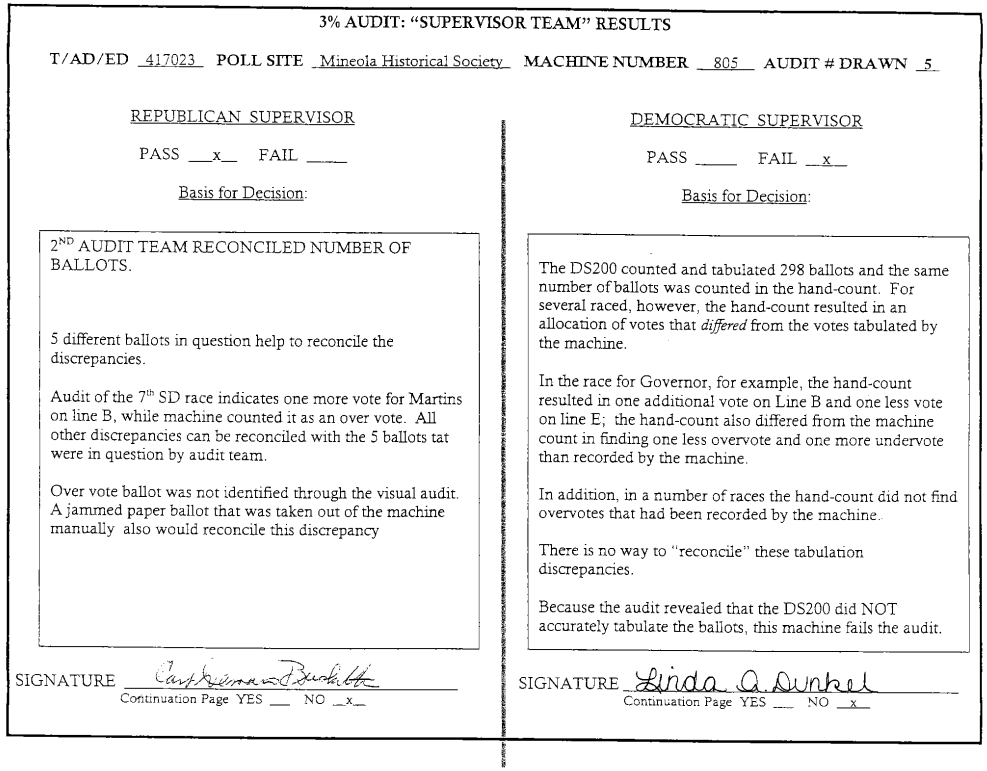
NY SD 7: Audit

- 7 of 249 of machines audited at random (3%).
- 3 of 7 (i.e., 43%) showed errors. Net error favored the apparent winner.
- Republicans: the errors were “reconciled”: Machines are fine. Democrats disagree.
- Judge Warshawsky: “In my opinion, reconcilable would be ‘Is there a clear reason why the deficiency occurred?’ ”
- Is “the machine was mis-programmed” a clear reason?
- In my opinion, reconcilable would be “We counted again by hand and found that the error was in the hand count, not the machine count.”
- “Clear reason” is irrelevant for whether the apparent outcome of the contest is correct. Size of the difference matters.

NY SD 7: Audit

- 7 of 249 of machines audited at random (3%).
- 3 of 7 (i.e., 43%) showed errors. Net error favored the apparent winner.
- Republicans: the errors were “reconciled”: Machines are fine. Democrats disagree.
- Judge Warshawsky: “In my opinion, reconcilable would be ‘Is there a clear reason why the deficiency occurred?’ ”
- Is “the machine was mis-programmed” a clear reason?
- In my opinion, reconcilable would be “We counted again by hand and found that the error was in the hand count, not the machine count.”
- “Clear reason” is irrelevant for whether the apparent outcome of the contest is correct. Size of the difference matters.

Example NY SD7 Audit report



3% AUDIT: "SUPERVISOR TEAM" RESULTS

T/AD/ED 417023 POLL SITE Mineola Historical Society MACHINE NUMBER 805 AUDIT # DRAWN 5

REPUBLICAN SUPERVISOR

PASS __x__ FAIL ____

Basis for Decision:

2ND AUDIT TEAM RECONCILED NUMBER OF BALLOTS.

5 different ballots in question help to reconcile the discrepancies.

Audit of the 7th SD race indicates one more vote for Martins on line B, while machine counted it as an over vote. All other discrepancies can be reconciled with the 5 ballots tat were in question by audit team.

Over vote ballot was not identified through the visual audit. A jammed paper ballot that was taken out of the machine manually also would reconcile this discrepancy

SIGNATURE [signature]

Continuation Page YES __ NO _x_

DEMOCRATIC SUPERVISOR

PASS ____ FAIL __x__

Basis for Decision:

The DS200 counted and tabulated 298 ballots and the same number of ballots was counted in the hand-count. For several raced, however, the hand-count resulted in an allocation of votes that *differed* from the votes tabulated by the machine.

In the race for Governor, for example, the hand-count resulted in one additional vote on Line B and one less vote on line E; the hand-count also differed from the machine count in finding one less overvote and one more undervote than recorded by the machine.

In addition, in a number of races the hand-count did not find overvotes that had been recorded by the machine.

There is no way to "reconcile" these tabulation discrepancies.

Because the audit revealed that the DS200 did NOT accurately tabulate the ballots, this machine fails the audit.

SIGNATURE Linda A. Dunkel

Continuation Page YES ___ NO _x_

HO 13-24 machine 952:

It was evident that the discrepancy between the audit (manual count) of the votes reflected on the ballots in the ballot box, and the votes reflected on the scanner result tape, are attributable to the two additional ballots found in the ballot box. The disparity between the manual count, and the scanner result tape, are precisely equal to the votes reflected when counting all off [sic] the ballots in the ballot box – including the two additional ballots. This is not "scanner error," but is instead attributable to any one of a number of alternative possibilities . . . The presence of more ballots in the ballot box does not demonstrate that the scanner has "failed," merely that the machine operated as it was designed to do – but with the result that some number (in this case, two) ballots were not scanned. . . . PASS

More

2-4059 Machine 104

1 additional ballot found and explains all discrepancies. . . . PASS

H18-12 Machine 259

One additional ballot was counted by the scanner than was found in the bin . . . PASS

Proffered testimony

The audit results not surprising even if a full hand count would show Mr. Johnson to be the winner.

Substantial possibility that the machine with the largest error was not one of the machines that was audited. 97% chance that auditing 7 of 249 machines won't check the machine with the largest error.

Average of less than two errors per machine could account for the apparent margin of about 450 votes.

Average of one error per 200 ballots could account for the apparent margin.

Proffer, contd.

Not a surprising level of error in precinct-count optically scanned ballots. Consistent with the errors the audit did find, within the statistical variability expected from “the luck of the draw.”

Large potential for error: the 242 unaudited machines could hold enough error to account for the apparent margin 186 times over. Sixty-six of the 242 unaudited machines could *individually* hold enough error to account for the apparent margin.

Proffer, contd.

Substantial chance that a 3% or 8% audit would find little or no error even if Sen. Johnson is the true winner.

If 30 of the 249 machines have errors of 15 votes or more—enough to account for the apparent margin—chance the 3% audit would have found any of those machines is under 60%.

If 20 of the 249 machines have errors of 23 votes or more–enough to account for the apparent margin–chance the 3% audit would have found any of those machines is under 45%.

If 20 of the 242 unaudited machines have errors of 23 votes or more (enough to account for the apparent margin) and an additional 5% of the machines are audited, chance the additional audit would find any of those 20 is under 69%.

Proffered testimony, contd.

Margin is so small compared to the possible errors that very large percentage of machines must be audited to give strong evidence that Mr. Martins is indeed the winner.

3% is not sufficient.

8% is not sufficient.

To have 90% statistical confidence that Mr. Martins won requires auditing a minimum of 90% of the machines selected randomly: an additional 218 machines.

This is true if the audit finds that those 218 machines have counted perfectly. If the audit of those 218 machines found many errors, still more machines would have to be audited.

What should an election audit law do?

Enunciate *principles*, not *methods*.

Methods are best left to regulation: Easier to improve, fix, etc.

Mutual distrust among election integrity advocates, elections officials, and legislators: unfortunate, makes good legislation harder.

California AB 2023 is an example of a good, risk-limiting law: `http://www.leginfo.ca.gov/pub/09-10/bill/asm/ab_2001-2050/ab_2023_bill_20100325_amended_asm_v98.html`

A Better Way: Risk-Limiting Audits

Risk-limiting audit: audit that has a guaranteed minimum chance of correcting a wrong outcome (by counting the whole audit trail).
Endorsed by CC, VV, LWV, CEIMN, ASA, . . .

Risk: maximum chance that the audit fails to correct an apparent outcome that is incorrect, no matter what caused the outcome to be incorrect.

Simultaneous risk-limiting audit: guaranteed minimum chance of correcting *all* the contests that have incorrect apparent outcomes.

Simultaneous risk: the maximum chance that the audit won't correct one or more of the apparent outcomes that are incorrect

Risk-limiting Audits in California

Marin County (February 2008; November 2008, 2009)

Yolo County (November 2008, 2009)

Santa Cruz County (November 2008)

Measures requiring super-majority, simple measures, multi-candidate contests, vote-for-*n* contests.

Contest sizes ranged from about 200 ballots to 121,000 ballots.

Counting burden ranged from 32 ballots to 7,000 ballots.

Cost per audited ballot ranged from nil to about $0.55.

Super-simple simultaneous audits

Truly simple: audit rules that allow elections officials to confirm that the outcomes of most contests are right, with one (small) sample.

Risk-limiting: large chance of correcting any outcomes that are wrong—i.e., that disagree with the outcome full hand count of the audit trail would show. (Correct them by conducting a full hand count.)

Exploit statistical efficiency of *ballot-level auditing*, which compares CVR with human interpretation of individual ballots.

Spend some efficiency to buy logistic and computational simplicity.

Sample size for ballot-level audits

"diluted" margin	risk limit 10%	5%	1%
5%	139	180	277
2%	346	450	691
1%	691	899	1382
0.5%	1382	1798	2764

With these sample sizes, can stop the audit if rate of overstatements is less than 20% of the reported margin.

Secret sauce

To implement ballot-level audits on a wide scale may require changes to vote tabulation systems: have to associate individual cast vote records (CVRs) with individual physical ballots.

Auditing using an unofficial vote tabulation system that does produce CVRs—such as those of Clear Ballot Group, the Humboldt Transparency Project, or TrueBallot—and confirming transitively that the apparent outcome is correct, might be the best interim option. (See Calendrino et al. 2007)

IRV and Ballot-Level Auditing

Even if precinct totals are perfect, cannot reconstruct RCV winner.

Ballot-level audits are (nearly) essential.

(Example by Rivest) Two scenarios, 5 ballots, 3 candidates (A, B, C). Same totals for each candidate for each rank, but different winners.

	Rank order	
Ballot	Scenario 1	Scenario 2
1	A B C	A C B
2	A C B	A C B
3	B A C	B A C
4	B C A	B A C
5	C A B	C B A
winner	A	B

What do we need for efficient audits?

- Laws that allow/require risk-limiting audits, but mostly . . .
- Data plumbing:
  Structured, small batch data export from VTSs.
  A way to associate individual CVRs with physical ballots.
- Reducing counting effort is mostly about reducing batch sizes.

What do we need for efficient audits?

- Laws that allow/require risk-limiting audits, but mostly . . .
- Data plumbing:
  Structured, small batch data export from VTSs.
  A way to associate individual CVRs with physical ballots.
- Reducing counting effort is mostly about reducing batch sizes.

What do we need for efficient audits?

- Laws that allow/require risk-limiting audits, but mostly . . .
- Data plumbing:
  Structured, small batch data export from VTSs.
  A way to associate individual CVRs with physical ballots.
- Reducing counting effort is mostly about reducing batch sizes.