CALTECH/MIT
VOTING TECHNOLOGY PROJECT

**A multi-disciplinary, collaborative project of
the California Institute of Technology – Pasadena, California 91125 and
the Massachusetts Institute of Technology – Cambridge, Massachusetts 02139**

WHO DOES BETTER WITH A BIG INTERFACE? IMPROVING VOTING PERFORMANCE OF READING FOR DISABLED VOTERS

Ted Selker
Media Arts & Sciences, MIT

Jonathan A. Goler
MIT

Lorin F. Wilde
MIT

Key words: *ballot interface, voter disability, experiment*

Who does better with a big interface? Improving Voting Performance of Reading Disabled Voters

Ted Selker[2], Jonathan A. Goler[2], Lorin F. Wilde[1]
[1]MIT Research Laboratory of Electronics
77 Massachusetts Avenue, Cambridge, Massachusetts 02139-4307
{jagoler@mit.edu}
[2]MIT/Caltech Voting Technology Project, MIT Media Lab,
20 Ames Street, Cambridge, Massachusetts 02139
{selker, wilde @media.mit.edu}
http://vote.caltech.edu/

Abstract: This study shows how ballot interfaces variably affect the voting performance of people with different abilities. An interface with all information viewable simultaneously might either help orient or overwhelm a voter, depending on his/her skill-set. Voters with diagnosed reading disabilities performed significantly better on full-faced voting machines than those who demonstrated a high likelihood of similar, but undiagnosed, disabilities. In contrast, the diagnosed group performed worse than others when using standard-sized Direct Recording Electronic (DRE) systems. We suspect that this observed difference in performance is due to the interaction of system features with learned coping techniques, which allow diagnosed reading disabled voters to function effectively in other parts of everyday life. The full-faced system provides a means of orienting but not of guiding the voter, while the standard DRE guides the users through the voting process without giving the voter a means of orienting themselves. A hybrid design that incorporates the advantages of both these systems might be beneficial for both reading disabled and non-reading disabled voters.

1 Introduction

Graphical user interfaces seek to make interactive information accessible by helping users find the information they want to work with and then focusing their attention so they can succeed in an interaction[1,2,3]. Voting interfaces present special design challenges. When a voter confronts a ballot, the complexity of an election and its issues is reduced to a simple choice of labels, which must be recognized and selected. We call this act syntactic because the finding and selecting of a label is entirely separate from the meaning of the choice [4].

Ballot design is vitally important to voting. The Voting Technology Project[5] demonstrated that equipment problems deprived between 1.5 and 2 million voters of their ability to vote for president in 2000. Further analysis by Dan Keeting of the Washington Post reveals that ballot design flaws caused 80% to 90% of these problems. In fact, the Palm Beach County data[5] reveals that over 19,000 people overvoted for president in 2000, while only 609 ballots had hanging chads. The overvotes due to the butterfly are not as relevant to litigation as the chads in which evidence exists and some change could be asked for.

1.1 Current Electronic Voting User Interfaces

Current electronic voting machines take several forms. Punch-card machines, which separate the marked punch-card ballots from the ballot information, performed worst overall, with 2.5% errors in voting for president in 2000. Optical scan machines, which proved no user feedback to show voters whether they have voted correctly, came in next to last with 2.2% errors. However, this number was reduced to 1.8% when voters using this interface were permitted to check their ballot with a scanner. In the 2000 election, we could not separate various types of so-called Direct Record Electronic Voting Machines (DREs). Together, DREs performed poorly, with 2.2% of voters not making a selection for president. Only 1.5% of voters using the old lever machines experienced this problem.

We expect that varying factors of process and bad ballot-design are at fault. Three approaches evaluate these effects. Firstly, Georgia reduced its statewide residual count (i.e. uncounted presidential votes) from an abysmal 3.2%, the worst rate nationwide except for Illinois, in 2000 to .4% in 2004 [6]. This astounding improvement is almost certainly a product of getting rid of bad processes as much as problem-causing ballots and voting machines.

Secondly, testing of VTP's Low Error Voting Interface (LEVI) [7] shows that improved ballot design can hugely reduce the human error rates compared to standard DREs. In fact, this research confirms the findings of other studies [8], which demonstrate that using many voting systems people make choices adjacent to that which they intend or make no selection on .5 to 3% of selections.

The third approach to demonstrating the impact of ballot design and to pinpointing opportunities to reduce lost votes is to see how memory and learning differences affect performance with simple layout changes. Our first work in this area has been to compare full-faced voting machines such as the ES&S V2000 against currently available, 15-inch DREs such as the iVotronic. Various attempts have been made to show how ballot variations affect users of different educational backgrounds[9].

Lever machines and DREs show voters that they have made selections; they also prevent voters from making too many selections on any race. People make more mistakes when their voting machine does not give feedback and remove the opportunity to cast overvotes (more selections than allowed). These broad forensic evaluations might be refined in various ways.

Interaction design improvements have typically been developed for normal users [10]. As with other areas, the voting rights of those with special needs are now protected by legislation. This legislation is a mandate to accommodate these disabilities during the interaction design of voting machines. Improvements to audio voting are already in progress [11]. This paper focuses upon visual interfaces in order to understand the implications of various ballot designs on the learning disabled community.

1.2 Learning Disabled Voters

Learning Disabilities is a generic term that refers to a heterogeneous group of disorders manifested by significant difficulties in the acquisition and use of listening, speaking, writing, reasoning, spelling, or mathematical abilities.[9] These disorders are intrinsic to the individual and presumed to be caused by

central nervous system dysfunction. Even such a disability may occur concomitantly with other handicapping conditions (i.e., sensory impairment, mental retardation), social and emotional disturbances or environmental influences (i.e., cultural differences, insufficient / inappropriate instruction, psychogenic factors) it is not the direct result of those conditions or influences. While the expected incidence of diagnosed and undiagnosed learning disabilities in the general adult population vary greatly, as many as 10-15% of adults in the workplace have been estimated to have learning disabilities [12]. Statistics as high as "30-50% of the population have undiagnosed learning disability" have been reported [13] based on sources from the National Institute of Literacy.

Difficulties with basic reading and language skills are the most common learning disabilities: as many as 80% of individuals with learning disabilities have problems in reading. [14]
In focusing this first study on comparative voting performance of reading disabled voters, we hypothesized that as much as ¼ of the general voting population could be impacted. However, no one has previously evaluated which characteristics of voting machines might improve the performance of this group.
Therefore, we sought to discover how particular technologies affect the ability to correctly mark selections for reading disabled and non-reading disabled voters. The promise of electronic voting machines is not only the ability to accurately and repeatedly count votes, but also to provide a means of effectively making selecting easier, flagging or disallowing voting mistakes and empowering populations that were previously disenfranchised from the voting process.

1.3 Historical Development of Voting Equipment
Lever voting machines provide a lever to pull with a finger for each selection in every race on a large panel. The levers are grouped so that rows with paper labels to the side represent political parties and columns with paper labels below them represent selections. Lever machines introduced the feature of presenting all races and selections simultaneous on a voting machine in the late 1800s to reduce some kinds of fraud. Such lever machines have been widely used for a century. They provide visual and tactile feedback to the users who could see and feel their vote as it is cast. Possibly because these machines stop people from voting for more selections than are allowed (overvote) and make it clear when votes are cast, they exhibit a low number of non-voted races at the top of the ballot [5]. However these outmoded machines have many administration and security problems that more modern solutions avoid. Voting administrators say that people do not vote as many races with them as with other voting machines. They could not provide alert-style feedback for common errors, such as inadvertently voting on less than the desired number of races (undervote). The older full-faced machines did not provide any form of record of actual ballots and could not provide correct count if in the event of common odometer failure.
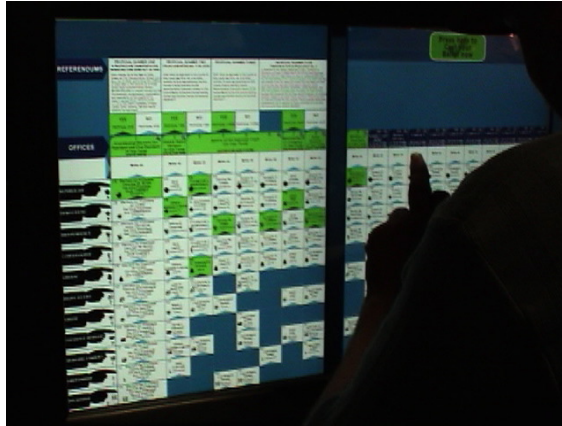
Starting in the 1860's the first attempts at electronic voting placed buttons behind or adjacent a paper or projected ballot. The introduction of commercial Direct Recording Electronic systems (so called because the recorded a vote without transcription onto paper) came with the Video Voter in 1976. More recent incarnations of direct recording electronic systems such as the Accuvote, dispense with requiring a paper ballot to be aligned with buttons and use a display and often include touch screens. The full-faced type systems provide the orientational advantages of lever machines. While the touch screens with a few races per screen guide the voter through the process. One of the great advantages of electronic voting is that they enable voting administrators to configure and deploy a ballot with far more ease than traditional equipment. Furthermore, the user interface, if separated from the vote tabulation modules, can be separately updated, allowing the latest in UI research to be applied.
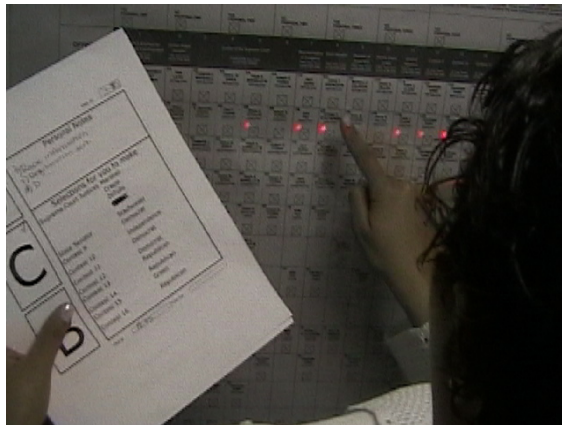

2. Materials and Methods

2.1 Equipment:

We acquired voting machines that are certified in New York and others that could be made available to NY voters. We contracted a YMCA that is used as a polling place in New York. Erie county 2000 presidential election was chosen to have many selections some familiar and some not. We hired people who normally

work as poll workers in New York and paid them what they normally get as poll workers. We also acquired voting educational materials from a NY county that the voters had not seen. We used Video cameras that were not intrusive (behind the head) to record votes off of voting machines. Although other voting companies were contacted only Electronic Systems and Software were willing to allow us to use their voting machines in our voting study. The voting machines we used were made available to us from Election Systems and Software. They included the currently available iVotronic DRE System with a 15 inch color touch screen Liquid Crystal Display (LCD) display, the V2000 Full Faced Electronic System and the prototype iVotronic LS Full Faced DRE with a massive high-resolution display.

Fig. 1. Voter voting on the LS Machine

Fig. 2. V2000

Fig. 3. Example Palm Card

2.2 Subject Selection:

The subjects of interest (reading disabled) RD subjects were located by advertising and professional contacts. Potential subjects were contacted by organizations to which they belonged to inform them of the study and to invite them to participate. For example, flyers were widely distributed to organizations, such as to local offices of the International Learning Disabilities Association, as well as to local unions and universities. In addition, we placed an ad in the New York Post to obtain both reading disabled and control subjects. At no time did the NYVS simply contact potential subjects directly. All subjects were to receive the same honoraria ($25/ hour for up to 2 hours).

As the RD subjects were differentiated from the non-LD subjects through the experimental procedure, it was not necessary to inform the individuals that they had been selected because they are or are not learning disabled. That is, at no time were subjects told that they were in an experimental vs. control group.

All subjects were required to be registered voters in New York.

2.3 Testing Procedure:

The voting experience was designed to simulate the emotional and physical experience of normal voting. Our goal was to have the voters learn about voting the way they normally do. We included a registration check in procedure, a typical length line to wait in for the tasks, some intervening tasks before using a second machine, and a check out procedure. The polling place was divided into three sections. The first section was the Greet and Education (GE) section. This section had a desk with two workers. Behind that desk was another set of tables with the reading test. The rest of the area was set up as a waiting and learning area, with tables and reading materials (newspaper with information about Erie county voting in 2000 and voter guides) associated with the voting process and the election. Beverages and small snacks were provided to and people were available to make sure that the voters would learn about the election before voting. The voters spent between 20 minutes and one hour there.

The voting section had 4 voting machines. The machines were each set up as they would in an election, with sufficient space surrounding the machines for privacy of voting. There were chairs for voters queuing to use machines. The final section was a post-test area. The post-test area had the post-test desk.

The testing procedure was to be as close to an actual voting situation as possible. We intended to recreate the physical characteristics of a polling station, but with a few modifications. First, we asked the voters to vote on the 2000 Erie County ballot with 13 defined races. We chose this election for the primary reason that New York City voters are unlikely to have voted on this particular ballot and were not be tainted by prior experience with the candidates/issues. Poll workers were contacted through the League of Women voters and were actual poll workers from previous elections.

When they arrived at the polling location, the voters checked in, and the poll workers explained the procedure for the experience:
1. Why we are conducting the study: We want to know how different types of people perform on different types of voting equipment.
2. What we are asking of them:
   a. To learn a bit about the candidates and issues that they will be voting on so that they can make an informed decision.
   b. To take a quick 5-minute reading comprehension test.
   c. Vote on two voting machines.
   d. Complete a short questionnaire about the experience
   e. Collect their honorarium.

The poll worker asked the voter to sign the informed consent form, and went over the form with the voter should they wish. Once the form was signed, the poll worker gave the user a copy of the form, which they reminded the voter contains the identification, and address of the researchers. They were then given a card with selections for the first 7 races and three letters on it indicating which stations they should go to when. The second station in each case was the reading and intelligence test. These numbers were generated from a random permutation of the numbers 1,2,3, and 4. This random element is part of the Latin-squares method of statistical analysis that we used. They were shown into the learning center, which had voter guides for them to learn from.

The voter was shown how to use the voting machine as a normal voter using the equipment would by an experienced poll worker and taken to the machine. Once the voter completed their first ballot, they were given the reading test. We then administered the 20-minute test assessing the RD/non-RD status of the voter. The test put after the first use of a voting machine was intended to distract the subject from the task of voting that they just completed. We believed this would give us better value for the results of the second voting experience. Once the reading test was completed, the subject was be given a break to get coffee/water, and then voted on the second machine. After their second vote, the voter was interviewed about their experience and given their honorarium.

3. Results

RD denotes learning disabled (with diagnosis) (n=41)
RD Control denotes control subjects who we determined were learning disabled (n=18)
Control are all control subjects (n=56)
MC is the upper IQ range of subjects who's average IQ matched the RD group. (n=18)

Table 1. Errors per ballot, by machine and subject classification

Machine	RD documented	RD undocumented	Control	MC
LS	3.3	5	4.55	3.4
V2000	2.5	3.4	3.9	4.5
iVotronic	6.45	4.3	5	4.7

Table 2. Mean time to complete ballot

Machine	RD documented	RD undocumented	Control	MC
LS	6:56	6:16	4:12	4:56
V2000	5:30	4:16	4:57	5:21
iVotronic	7:00	7:10	4:22	4:00

Table 3. Races left blank per ballot

Machine	RD Documented	RD undocumented	Control	MC
LS	4	4.08	4.8	2.8
V2000	2.76	3	5.43	4.25
IVotronic	1.9	1.27	1.64	0.2

Table 4. Self-reported discomfort, scale of 1-5 with 1 being the most uncomfortable

Machine	RD documented	RD undocumented	Control
LS	3	1.86	2.25
V2000	2.91	1.8	2.23
IVotronic	2.5	2	2.5

Table 5. Errors per Race Completed

Machine	LD	LD Control	Control	NDC
LS	0.18	0.26	0.25	0.16
V2000	0.13	0.18	0.23	0.26
iVotronic	0.29	0.19	0.23	0.21

3. Discussion

Diagnosed Reading disabled voters achieved significantly lower error rates for the top races on full-faced machines than they did on the standard DRE machine: 3.3, 2,5 vs. 6.5 for LS, V2000 and iVotronic, respectively [Table 1]. Our hypothesis is that the lower capability "control" group and the undocumented reading disability group did not benefit from sort of coping system that the documented reading disabilities people had developed. They performance on these full faced machines is interesting in that they got almost ½ the errors on them that they did on the DRE machine. The undiagnosed RD voters exhibited a higher error rate (5 vs. 4.3, LS vs. iVotronic), showing that they found the full-faced systems a bit more difficult to vote on.

Table 2 shows how the documented reading disabled people took longer (confidence t< 0.01) and than the control or matched control subjects and did as well or better error wise than them. Interestingly the profile of how long the undocumented reading disability people took was similar but did not "benefit them" in the same way.

Table 3 shows that the races left blank per ballot were lowest in every group using the DRE touch screen. Many people have pointed out that the structured voting experience small electronic displays put forward of taking voters through the races reduces reduced voting lower on the ballot. Indeed such down ballot falloff was reduced for all voters in the study. We also expected that the errors would have been reduced across voters. We believe that the ability to easily review the ballot on the full faced in this study may have been responsible for the reduced error. We regret that we weren't able to introduce the LEVI voting system into this experiment. The LEVI DRE has been shown to reduce voting error rate significantly in such DRE interfaces. If it reduced the error rate by the amount it has in our other experiments it would have eliminated this DRE problem.

Undocumented reading disabled voters clearly had more problems than the other groups, Table 4 shows that they appropriately found all machines less comfortable to use.

One of the more striking differences between subject groups is the level of self-reported discomfort using a particular voting system, as noted on the post-test questionnaire. While voters never actually can see their vote being stored, the self-reported discomfort is more likely to affect their stated decision to participate in voting in the future. Voters who were learning disabled and did not know it were uncomfortable voting, but the voters who had diagnosed learning disabilities had significantly less discomfort when voting on full-faced machines.

In this study the control group exhibited poor performance on all systems. We were concerned about the poor performance on all the systems for the standard control subjects. Based on the (name of test?) estimated IQ test, the average IQ of controls was significantly lower than the average IQ for learning disabled voters. We suspected that a difference in IQ might have confounded our results for performance. We therefore broke out the higher IQ control subjects into a group whose average IQ matched that of the diagnosed RD (n=34) . Matched Control MC subjects exhibited the lowest error rate on the LS machine, and MC subjects had significantly less blank races on all machines when compared to normal controls (Table 3).

Finally, we are concerned with the high error rate on the ballots. With the average residual rate for some group/machine combinations ranging from 5 as high as 9 races per ballot, we must seriously look at two factors: the voters are either incapable of accurately recording the voter intent, or the voters ignored or were unable to use our palm cards effectively. We had the advantage of analyzing the data using video capture of the voters filling out ballots and casting their votes. Thus, we were able to watch the voters fill out ballots, realize errors, and react to the machines being unresponsive. In the entire study, only one voter reacted violently to the machine (the prototype LS) when it was hanging. In a few cases, a voter went through the entire ballot without looking at the palm card, selecting the same row of candidates for every race. We disqualified those ballots from our analysis because it was clear that those voters were not complying instructions allowing us to test errors with the experiment. We are convinced that the differential in error rates does indicate a true difference in a voter's ability to successfully and accurately vote.

It is clear however, that voters need more education in how to use all systems before they are permitted to use them. For all subjects, page-by-page DRE systems significantly reduce undervotes, while increasing the total time to complete the ballot. Clearly, by forcing the user to take action on each screen encourages them to vote on other issues. This study does not establish whether or not those voters make intentional or random decisions on races that the DRE 'encourages' them to vote on. For most precincts, the full-faced system appears to provide a significant advantage for completion time. This gave them higher throughput of voters.

4. Conclusions

Interaction design in graphical user interfaces has seen much improvement over the years. Interfaces have progressed from blinking text to smooth scrolling text to using modeling, shadows, grouping, and metaphor to improve interaction. In our study, 13 races had named selections but even in the best case, voters made 2 or more errors. In voting, the stakes are high and one error in 5 selections seems atrocious. While accuracy means a lot in command and control interfaces, they have feedback, semantic connection and are highly practiced.

Voting is not frequently practiced. Voters do not vote often and their selections are hopefully varied throughout the years. We believe that this requirement for accuracy in rarely used interfaces may not have been adequately studied for other interfaces. In this rarefied problem of syntactic interface accuracy, we can find new problems and document how orienting and focus prosthetics may make large improvements.

One expects that reading disabled people might have more problems with a full-faced voting machine: lever machines have an overwhelming number (up to 509) of active switches.

The findings are much more interesting considering this daunting number of choices. The study shows that full-faced voting machines are faster and more accurate if the users are organized and careful in using them. This result comes out in several ways in this study. The low IQ subjects who have problems of organization and follow-through have the most errors with unguided voting. For tested reading disabled people, we found them using coping mechanisms and outperforming others for errors on the full-faced voting machines. Interestingly they did not do as well with the structured DRE experience. Could this be because they could not move fluidly around the voting ballot, checking their work? If so LEVI will help.

The differences between groups are significant. If you are diagnosed as reading disabled you seem to be a more careful voter. This difference may be because you learn and develop compensatory behavior. For example, in our video recordings, we observed RD voters using their fingers or paper to line up the rows that they were voting on, and double-checking their answers.

Clearly, different technologies have benefits for certain groups of individuals. Full-faced ballots are more effective for all groups except for non-diagnosed learning-disabled voters. The intuitive nature of a full-faced ballot is easier to vote on for most non-RD subjects and allows them to complete their ballots more quickly and with fewer errors. In addition, diagnosed RD subjects are able to use the grid structure of the full -faced ballot to orient themselves when voting, while they do not have this orientational benefit on the page-by-page DRE.

This paper demonstrates that the cognitive prosthetics of structured experience with a DRE are especially helpful to some and make everyone vote more. Sociologists have employed other cognitive prosthetics such as encouraging people when they get part way through a task [nass west02 talk]. In this experiment we note that the full-face orientation augmentation may have given the careful organized voter the chance to scan carefully for errors.

People use interface and human prosthetics together to perform in any task. Voting interface research can be at the forefront of seeking understanding for how to create orienting interfaces that guide people to compete all entries and check them carefully. So far, different interfaces have different trade offs, research projects should strive to create interface technologies that universally help all users. The LEVI project seeks such goals.

The protocol developed for this research allowed us to realistically run an election experiment across 3 machines with 3 groups in one day. The protocol was improved in stages and through many changes. At first we thought that we should rely only on documentation for reading disabilities. This would have lead to missing most of the results of the study. We wouldn't have found the undocumented reading disabilities people in the group. Second we relied on Larry Thomas' assessments to test for IQ. We did not expect what we found, that the control group would be differentiated by their lack of cognitive skills. Third we found that the change of putting the tests between the experimental conditions made many improvements to the study. First it evened out the flow of subjects. Second it changed the pace and we are convinced made the second cell much more useful in our study. We also found that video taping the voting machine was extremely useful, in the first place it gave us images of each individual ballot, it also allowed us to assure ourselves of difficulties and coping mechanisms that voters used while they were voting.

We are concerned that some of the voting officials we used were not well trained or suited for the job. In one case, a worker who reported having worked on numerous elections was obviously failing at comparing names said to those on the registration book. In this case we finally relieved them of their responsibilities at 10:00 and checked their IQ using the same tests as the other subjects. It shocked us how difficult it was to get poll workers that would follow instructions and complete tasks. Another poll worker tried to aid voters, while another sat still and ignored requests for help. In the future we will include more poll workers. The sign in, pretest, learn, vote, test vote, post test protocol was extremely effective.

5. Future Work

We plan to use the protocol developed for this study to test various new user interfaces aimed at taking the orientation-aiding advantages of the full-faced systems and combining them with the guided nature of page-by-page DREs. Hopefully the resulting system will be less confusing for voters and permit them to accurately record their intent.

We will be integrating the results of this study into the LEVI system, which will demonstrate that it is possible to improve voting performance for many groups of people. Already the LEVI system includes a visual interface that provides both a full-view of the ballot as well as guiding the voter through the ballot. By doing both, we will provide the orientational benefits of a full face machine, while also preventing the voter from unintentionally leaving races blank.

Acknowledgements

References

1. B. Schneiderman. *Direct Manipulation: A Step Beyond Programming Languages*. IEEE Computer, 16(8):57--69,1

2. Roth, Susan King, Disenfranchised by Design: Voting Systems and the Election Process, Vol. 9, Number 1. Information Design Journal (1998) 29-38) http://www.informationdesign.org/pubs/roth1998.html

3. Ark, Wendy, Dryer, D. Christopher, Selker, Ted, Zhai, Shumin: Representation Matters: The Effect of 3D Objects and a Spatial Metaphor in a Graphical User Interface. People and Computers XIII, Proceedings of HCI '98, (eds) H. Johnson, N. Lawrence, C. Roast. Springer (1998) 209-219

4. Klatsky, Roberta: Human Memory Structures and Processes. W.H. Freeman, San Francisco (1980)

5. Alvarez, Michael, Ansolabehere, S., Antonsson, Eric, Bruck, Suki, Graves, Steve, Palfrey, Tom, Rivest, Ron, Selker, Ted, Slocum, Alex, Stewart, Charles S.: Voting, What Was, What Will Be. (July 2001) http://www.vote.caltech.edu

6. http://mit.edu/cstewart/www/

7. LEVI – Unpublished MIT/Caltech Voting Technology Project

8. S. M. Sled, Vertical Proximity Effects in Recall (2004;http://www.vote.caltech.edu/Reports/vtp_WP6r.pdf).

9. Tomz, Michael, Van Houweling, Robert: How Does Voting Equipment Affect the Racial Gap in Voided Ballots? American Journal of Political Science Vol. 47, No. 1 (January 20003) 46-60

10. Hammil, Leigh, McNutt, and Larson: (1981, 1987) 109

11. http://vote.caltech.edu/Reports/selker.ppt

12. http://www.ldonline.org/ld_indepth/adult/ld_and_special_learning_needs.html

13. http://www.floridatechnet.org/bridges/factsandstats.pdf

14. http://www.twc.state.tx.us/svcs/adultlit/litdisability.html

15. http://www.vote.caltech.edu/west02/pdfs/Nass-C.pdf