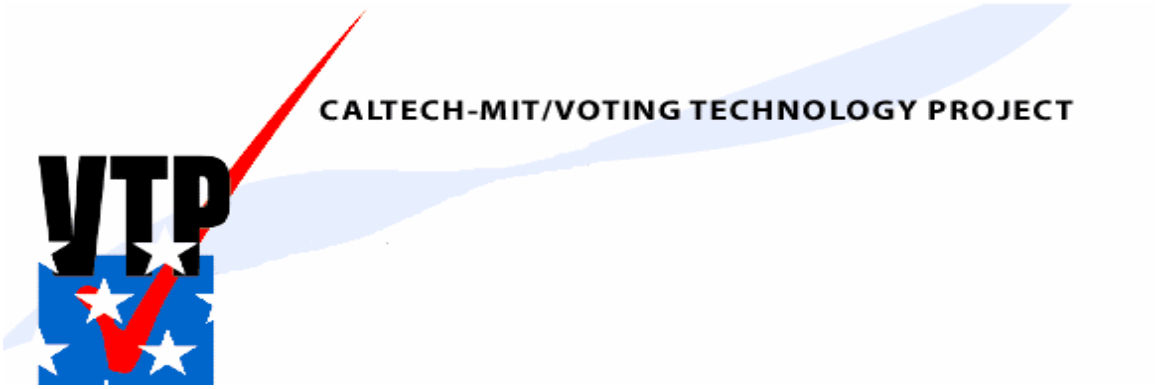

Qualitative Evaluation: "VotoElectronico: Prueba Piloto 2005, Ciudad De Buenos Aires"

R. Michael Alvarez, Ph.D.
Professor of Political Science
Caltech/MIT Voting Technology Project
California Institute of Technology

November 22, 2005

Executive summary

The Buenos Aires "VotoElectronico" pilot project took place on October 23, 2005, in the City of Buenos Aires. It involved a pilot test of four different electronic voting systems, in at least 43 voting stations located throughout the City. The purpose of this report is to provide initial qualitative assessments of the pilot project. Quantitative analysis will hopefully be presented at a later date, once detailed data from the pilot project are available for examination.

The key recommendations for the current pilot project evaluation effort and for the possible future continuation of this pilot project, made in this report are:

1. **Much care is needed in analysis and study of the evaluation data.**
2. **Determine the extent of usability problems.**
3. **Take a careful look at the optical scanning system.**
4. **Develop and implement procedures for physical security.**
5. **Develop and implement procedures for pilot testing security.**
6. **Continue pilot testing, but narrow down the number of options being tested.**

Below, I discuss the pilot project in more detail, provide a detailed discussion of my observations, and provide more discussion of these six key recommendations. Further details from the project and from my observations are available at http://vote.caltech.edu/Elections2005/BuenosAires/buenos-aires.

Project background

The "VotoElectronico" project was the culmination of a research effort that is described in background materials provided to the international observation team, and those materials are available at the Office of Electoral Affairs, City of Buenos Aires (http://www.buenosaires.gov.ar/dgelec). The international observation team (comprised of a large group of observers from across the world, including both public and private organizations) participated in a series of meetings on October 21, 2005, including the opportunity to examine each of the voting systems that were part of the pilot project. On October 22, the observation team was able to observe the set-up of the electronic pilot project systems in two different voting sites. Last, on October 23 the observing team split into two groups, and each group was able to observe both normal polling place observations as well as pilot testing at a small number of poll sites located throughout the City of Buenos Aires.

During the implementation of the pilot project on October 23, it appears that a great quantity of information was collected from project participants: a short questionnaire that was given to all participants; a longer questionnaire that was given to a subset of participants; data on voting and system performance from the voting systems themselves; and qualitative information from the polling place workers participating in the pilot project. This more detailed data will hopefully be available once it is compiled and ready for study; at this point, the data at hand are based on qualitative observations made during the voting system comparison by the observing team on October 22, observations made during the pilot test on October 23, and four graphs of participation data that were received immediately after the pilot test via email. Additional information on the pilot project, and the observations of Sean Greene and Dan Seligson of electionline.org, can be found in the October 27, 2005 edition of "electionline Weekly".[1]

Description of the voting systems

The pilot project was quite ambitious, involving the use of four quite different voting systems. Currently, voters in Argentina use a party-list paper ballot system when they vote. Thus, the voting systems tested here all were developed with the existing institutional process of Argentina in mind.

The four different voting systems were developed and built using an innovative public/private approach; the open-source software was developed by the public sector (within the Office of Electoral Affairs), with the hardware provided by private sector firms selected after a bidding process. The voting system hardware was engineered to fit inside a relatively large black metal casing, each voting system thus being approximately four feet high. Examination of the inside of each of the voting devices revealed that each involved the integration of a standard off-the-shelf computer system, connected to a variety of input devices (the input device thus being the primary source of variation

[1] http://www.electionline.org/Newsletters/tabid/87/ctl/Detail/mid/643/xmid/160/xmfid/3/Default.aspx.

across the voting devices). As a result, by employing the different voting systems using relatively similar external casings, the pilot test team attempted to keep constant any technological variables that might have influenced voters participating in the test.

It is important to stress that the four voting systems tested on October 23 are "concept" systems --- they are not necessarily designed nor engineered as they might be were they used in future elections in Buenos Aires or Argentina. This pilot test, then can be seen as a large-scale "proof of concept" test, involving four different concepts for how voters in Buenos Aires (and possibly Argentina) might be able to cast ballots in future elections. Further details of the voting systems, including demonstration material provided to the observation team, can be found at http://vote.caltech.edu/Elections2005/BuenosAires/buenos-aires-docs.

The four different voting systems, as used in the October 23 election, were numbered (and provided acronyms) and these numbers will be used in this report to refer to specific voting systems. Voting system 1, which was referred to as the "REA" voting system, involved a two-stage process involving the use of a plastic "smart card". Voters would insert the smart card into a voting system that had a computer screen and a touchpad. They would use the keypad to navigate through screens to cast their ballots; when done, they removed their smart card, moved to a second machine, and again inserted their smart card --- this recorded their vote.

Voting system 2, which the pilot testing team referred to as the "REV" voting system, was also an electronic voting device. However, instead of navigating through the ballot with a keypad, voters using this system used a touchscreen to cast their ballot after activating the system with their plastic smart card. When done with their ballot, a paper audit trail would be generated underneath a glass screen. If the voter affirmed that that indeed was how she wanted her vote to be cast, the paper audit trail fell into a bin and the voter was done; if not, the paper audit trail was rejected and the voter was allowed to cast the ballot again.

Voting system 3, which was called "LOB", was also an electronic voting system, though quite different from the previous two. Here, the voter did not use a plastic smart card to active the system and obtain the correct ballots. Instead, inside the voting booth, the voter picked paper ballots for the party list she wished to support inside the booth, and inserted those ballots into the machine. The ballots were scanned, and the voter was provided a number of opportunities to confirm or reject the party list selection. After a final confirmation screen, the voter's ballots were accepted and the voter was finished.

Voting system 4, termed "LOP" by the pilot testing team, was an optical scanning voting device. The voter obtained a large paper ballot that contained a list of all the parties running candidates for office in each of the two races on the ballot (this lead the ballot to be on two very large pieces of paper, connected in the middle with a flexible binding). The voter used a special marker to make her selections then folded the ballot so that the selections for each race faced each other. She then inserted the folded ballot (again, folded so that her selections could not be seen by others) into the voting device, where it

was scanned for mistakes. If a mistake was detected, an error warning was triggered and the poll site worker was to provide another ballot or allow the voter to fix the existing ballot; if no mistakes were detected, the ballot was accepted and fell into a ballot box.

Implementation of the pilot test

The pilot test team selected 53 polling locations throughout the City of Buenos Aires to participate in the pilot test. Based on a visual inspection of a map of the City and the location of the pilot test sites, it appears that they attempted to locate a pilot test site in each part of the City. According to a briefing provided to the observing team, during the two weeks prior to October 23, an information campaign was mounted to contact voters and inform them about the pilot project; we were told that this information campaign involved letters to voters in pilot test areas, street signs, and attempts to get the local media to cover the pilot test. However, we were informed that in at least one voting site that we visited, for some reason letters were not sent to eligible voters. Also, on Election Day we were told that pilot test sites were not operational in at least 9 of the 53 test sites, thus possibly reducing the set of pilot test sites to 44. A histogram of participation rates in each pilot test site provided to me via email shows data for only 43 test sites, thus indicating that the pilot test might have been conducted in ten fewer sites than originally planned. After more detailed data is provided for the pilot test, it will be possible to more accurately ascertain the representativeness of the pilot test sites as planned, and then as implemented.

On October 22, the entire international observation team went to two pilot sites to observe the set-up of the pilot test site. These two sites were both schools, one located at Av. Lope de Vega 2150 ("ENET No. 35 ING EDUARDO LATZINA"), the second located at Av. Santa Fe 2778 ("LIC NAC No.1 JOSE FIGUUEROA ALCORTA"). On October 23 (Election Day), the observation team was split into two groups; the group that I was with went to five different pilot test poll sites:

- 32 "SOCIEDAD RURAL ARGENTINA", Av. Santa Fe 4201.
- 38 "ESC No. 25 BANDERA ARG.-EMEM No.6 Padre C. MUJICA", PJE EEUU Del Brasil y Ave. ANT.ARGENTINA.
- 37 "ESC No. 6 FRENCH Y BERUTI", Basavilbaso 1295.
- 32 (again).
- 42 "ESC No. 21 SOLDADOS ED MALVINAS", Goleta S. Cruz 6999 (Gral.PAZ y 2 de ABRIL).
- 20 "ESC CANGALLO SCHULE", TTE. Gral. Jan Domingo Peron 2169.

The basic plan of action in each pilot test site was for the site to be set up in an easily accessible location, and for some of the pilot test workers to approach voters as they came or left the poll site, hand them a brochure, and ask them if they would like to participate in the pilot test. If a voter agreed to participate, they were directed to a registration table, where their identification card was checked, and if eligible, they were then allocated randomly to one of the four test voting systems. If the voter needed a smart card to operate the voting system they were selected for, they were given that smart

card at that point. The voters were then directed to an information booth for their voting system, which had another pilot project worker inside the booth who walked the voter through a set of instructions provided on a poster on the wall of the voting booth. Once the voter completed the orientation/information session, they were sent to another voting booth where they were given the opportunity to cast a mock ballot for the October 23 election. Upon completion of their mock vote, most voters completed a very short questionnaire; a randomly selected sample of voters completed a longer questionnaire assisted by another poll site worker. After completion of the survey, the voter returned to the registration table where they received their identification materials and then left the pilot test site.

Before the pilot test, the test team stated that they hoped to get as many as 20,000 participants in the pilot test, distributed throughout the 53 pilot test locations in the city. However, we were told late in the evening of the election that at least nine of the pilot test sites were not in operation on Election Day, apparently having not been allowed to operate by uncooperative poll site judges. But we were also told that evening that as many as 14,000 may have participated in the pilot test.

There are some obvious caveats to make at this point regarding this pilot test. While highly ambitious and well designed, there are some potential issues to note at this stage before any data are actually analyzed for more detailed project evaluation:

1. This is a very ambitious pilot test, involving four vastly different voting systems. As described above, they differ in many ways, including the basic way in which the voter activates the device and receives the correct party-list ballots, how voters select their choices and interact with each device, how voters confirm and verify their choices, and what they need to do to conclude voting. Some of the voting devices require that the voter manipulate paper, others are completely electronic. Some have keypads for information entry, one a touchscreen, and one had no direct data input device for a voter to use. Thus these four devices vary widely, on many dimensions, making it difficult or even impossible to be able to know with precision exactly what attribute of one voting system might lead it to have superior performance over another in evaluation testing.

2. According to materials provided to the observing mission, the criteria for pilot test site selection are based on a variety of demographic and prior electoral participation variables; at this point it is not clear how representative the pilot test sites are of the overall population of voting sites in this election throughout the City of Buenos Aires.

3. Whether the pilot test sites that were not operational on Election Day were somehow systematically different from those that were operational is not clear at this point, and could also be a source of potential selection bias.

4. Subjects self-selected to participate, and thus there might be selection bias in the basic evaluation data. Also, the pilot test workers (especially those observed working in the pilot test sites we attended) were all very young and this might serve to influence the type of participant who desired to try the pilot voting project.

5. Participants were not casting real ballots, instead were casting mock ballots. Thus it is not clear whether voters necessarily had a strong incentive to pay attention during the information session, to avoid errors when they voted, and to respond accurately to the follow-up surveys. These are unavoidable issues in mock election pilot tests.

6. The potential effects on the pilot test evaluation of the failure of the voter education campaign (the observing team was told that the mailings to eligible voters in at least one of the pilot site locations were not sent) should be assessed and factored into the evaluation. Also, pilot test workers were observed possibly "updating" their voter training as day progressed, and this is a factor that should be taken into account (if possible) in evaluation. Last, the observation team went to at least one test site that was understaffed (we were told that of the three workers who were supposed to assist with participant recruitment, only one arrived; while we were in that pilot test site, that worker was off in another location casting his or her own vote, thus leaving the pilot site without anyone actively recruiting participants while the observation team was there).

These issues do not necessarily make the pilot test meaningless, though paying careful attention to testing for potential biases and for employing methods from modern statistics to alleviate these potential biases are essential for a meaningful and productive final evaluation.

Important observations about usability

During the hands-on evaluation by the observation team on October 22, and in my observations of voters using the four voting systems on election day, a number of usability issues became clear, some specific to certain voting systems, other more general issues. Of course, as these voting systems are in the conceptual phase, my hope is that these observations will help the pilot test team improve these voting systems or focus their efforts on a more limited set of voting systems in future tests.

First, and of most importance, there were clear problems with voting system "4", the "LOB" optical scanning device. Without a more detailed examination of the mechanical details of this device, it is hard to determine precisely what the problem was --- usability, mechanical problems, computer issues, or some combination of these and other factors. Members of the observation team themselves had difficulty using this voting device, as during our hands-on testing and during the pilot test on October 23, we observed:

- Voters having difficulty inserting the ballot into the device.
- The device frequently had difficulty accepting a correctly-completed ballot.
- The device was frequently out of operation on October 23.

Again, the observation team did not have an opportunity to examine the mechanical and electronic operations of this device in detail. One feature of this voting system might merit future examination; the optical scan ballot that was used involved heavy stock paper, a special marking pen, and the need to fold the voted ballot on top of itself for privacy. These features of the paper optical ballot might make a voted ballot difficult to read by an optical device, and thus project team members might examine alternatives to this approach, including different optical scanning hardware and the use of different types of paper optical scanning ballots (especially the use of a "privacy sleeve" instead of a ballot that folds on top of itself).

Second, the two voting devices (systems "1" ["REA"] and "2" ["REV"]) that ask voters to interact with the computer screen to navigate their ballot might require some additional development. The specific issue here is how the ballot is rendered on the screen, and the sequential way in which voters are asked to work their way through the ballot. One qualitative issue that I observed from some voters on October 23 who used these devices was that some wanted to see the entire ballot rendered on one screen; one voter said something to the effect that he "wanted to see all of his choices in one place." Indeed, in party list elections it could be difficult to try to render all of this information in a usable way onto a small computer screen.

However, there are both hardware and software solutions to this issue that might be considered for further development of these two conceptual voting systems. A hardware solution (potentially expensive) would be to utilize a larger computer screen (especially for future development of the non-touchscreen systems), as this would allow for more screen space, and thus allow additional information to be presented to voters. A software solution might be to use other types of interfaces for the voter, including interfaces that developed by my MIT colleague Ted Selker and his research group, the "LEVI" interface; "LEVI" uses clever tools like a menu on the left of the screen, a "fish-eye" navigation approach that increased the font size of objects the voter is currently viewing larger than the objects not being viewed, and the use of various colors and fonts to indicate selections that the voter has made or has yet to make.[2]

Third, there is clearly more research needed as to how to employ "smart card" technology in future voting systems for Buenos Aires voters. Voters (especially elderly voters) were frequently observed having trouble trying to insert their smart card into each of the voting systems that used that technology. Interestingly, there were some voters who appeared to try each possible permutation of smart card insertion before getting the "correct" one, causing much frustration and slowing down their voting experience. This appeared to be a frequent question or problem for voters in the pilot test.

[2] See Selker et al., "Orienting Graphical User Interfaces Reduces Errors: The Low Error Voting Interface", Caltech/MIT Voting Technology Project, VTP Working Paper #23, February 2005, http://vote.caltech.edu/media/documents/wps/vtp_wp23.pdf.

The education posters for the voting systems using smart cards appeared to not have demonstrated the correct use of smart cards, and I did not observe pilot test workers demonstrating the use of the smart card to voters (this may have happened, I just did not observe it). Also, it appeared that voting system "1" ("REA") showed the smart card's insertion on the screen incorrectly, potentially confusing some voters.

Thus, future tests of voting systems involved smart card technologies should focus on a more usable process for voters, and should also involve some quick education for voters as to how to use the smart card. For example, having clear and accurate instructions on a future voting system about smart card insertion should help, as would having smart card instruction before a voter is allowed to use the voting system (for example, if this process moves to a point where a single voting system is used in a future pilot or election, voters could be walked through a tutorial by a poll site worker after having checked in, but before they are allowed to use a "live" voting device).

Fourth, the process used by voting system "1" ("REA") seemed unnecessarily cumbersome, and I could not ascertain why there were independent devices used for ballot generation and ballot recording. There are many potential drawbacks with this approach, including that by having more devices in a poll site there are additional points of failure and additional potential threats, and that by having multiple steps for a voter that they may have additional troubles. One concern is that some voters in a broader implementation may forget the second step, or if a line has formed in front of a ballot recording device they may want to skip waiting in line for ballot recording; if this were to happen, significant anomalies in the audit logs of the two devices would exist and it might be difficult to rectify why additional ballots appeared to be generated relative to the number recorded.

Fifth, and more generally, some additional research on usable poll site layout might be necessary. Some of the poll sites that the observing team visited had a well-designed layout; there was a clear and linear progression from the check-in point, to the instruction booths, to the voting devices, and back to the check-out point. In other poll sites, the layout was confusing and difficult to easily navigate (in some cases this was most likely due to the available space for the pilot test in each voting site). If possible, future pilot tests should more closely resemble the existing process used traditionally by Buenos Aires voters, with the same type of linear flow from check-in to ballot generation and back to check-out. Also, another aspect of poll site layout that was frequently neglected was how to appropriately orient the electrical and computer cables in the poll site; exposed cables that were neither well marked nor well anchored were observed in every poll site visited in the pre-election period and on Election Day. In some cases, the cabling was left exposed in high traffic areas, posing a physical risk for voters as well as a potential risk for security or voting disruptions.

Sixth, some issues arose with regards to the paper audit trail produced by voting system "2" ("REV"). One issue was that some voters clearly did not understand that the paper audit trail being produced behind the glass shield was not for them to handle. In some

pilot test sites voters were observed trying to access the paper audit trail by trying to remove the glass. In one pilot test site, a voter had actually removed the glass shield (or had broken it); later the observation team noticed that other voters were trying to access the paper audit trail as the glass shield was no longer in place. Clearly this usability issue should be resolved with better voter education, both before and during the voting experience.

Procedures and security

Observations were made regarding a number of aspects of voting system security, and there were a number of different issues that arose, before and during the pilot testing project.

Unlike the existing situation, where party-list paper ballots are produced, distributed, and counted, movement to a process involving electronic voting devices (both optical scanning and electronic voting machines) requires a transition from a largely paper-based process to one that is heavily dependent on computers and computer software. Accordingly, new procedures to insure the security of these computers and computer software should be developed and implemented before additional pilot testing is undertaken.

One set of security procedures that should be considered involves the physical security of the software development process. During the pre-election period, the observation team noted a relative lack of physical security in the office areas where we were told that software development for the pilot test had occurred. In those office areas, doors were open and unlocked, computers were left on, these computers appeared to be connected to some type of network or directly to the Internet, and there were windows in these offices that were left open. This was generally true of the election office, as far as the observation team could ascertain, that there was not a strong emphasis on physical or computer security. As this was not an important component of the observation team's activities, it is unknown whether this was an anomaly, or standard procedure.

Secondly, during the pre-election set-up period for the pilot testing poll site, there was a stronger emphasis placed on physical security, and that access appeared to be strictly limited to election officials, polling place workers, the observation team, and the police or security forces guarding each pilot test site. However, once within the site, during the poll site set-up the voting devices appeared to be unsealed and not supervised, nor was there a process that we observed that required the necessary records to be generated and retained to insure that an audit trail was being generated including information as to who had access to each component of the pilot test at each moment in time, and whether each manipulation required to set up the pilot testing equipment was being logged and supervised by multiple poll site workers or election officials. It was not clear that equipment for the pilot testing was being tracked, logged and supervised, at least by quick visual inspection in one of the pilot test sites being set up before the election.

Third, we witnessed a number of different issues associated with a lack of security procedures, or procedures that were not being followed. In one pilot site we noticed that the taped seals had been removed (or had not been installed) from at least two of the pilot test machines. We also observed in one pilot site one of the electronic voting machines (system "1", "REA") had stopped working; one of the pilot test workers appeared to reboot the system a number of times (unsupervised). I assume that these reboot attempts were logged by the system, but as best as I could tell there was no external record created as to the poll worker's manipulations of this voting system, nor were his actions supervised by any other poll worker that I noticed.

Fourth, in some pilot test locations the pilot test site's physical layout made it difficult if not impossible for either the pilot test site workers or outside observers to see the voter's actions on the voting machines (note that I am not asserting that the voter's choices must be observed, but instead that any other actions a voter might take may be difficult to observe). For example, without direct line-of-site visual observation of the voter's use of the voting system (again, done in a way that will preserve the voter's privacy) it might be difficult for poll site workers to see situations when voters are having obvious problems with the voting technologies. Clearly the physical layout of the polling places need some careful scrutiny in the future, trying to orient the voting technologies to maintain secrecy and privacy, while also insuring that the polling place workers can observe any problems with the voting systems or attempts to tamper with the voting systems.

However, the pilot test project also has made software running the electronic voting technologies available to the public and the observation team, thus adhering to an "open source code" ethic. This is a very important observation to make, and this project should be applauded for making their source code available for outside inspection and scrutiny. In some ways the open source approach may help mitigate future concerns about security of the voting technologies, and may lead to important improvements in the software itself if and when outside experts provide feedback about the code upon examination. I hope that the open source approach is maintained as this pilot project moves forward with system improvements towards the goal of actual system implementation.

Key recommendations

1. **Much care is needed in analysis and study of the evaluation data**. As noted above in this report, there are many questions about the representativeness of the data generated by this pilot test project, as well as other concerns about the reliability of the data. Very careful analysis of the pilot test data will be required to insure that potential biases are identified, and if possible, mitigated or accounted for in the final evaluation analysis. While the initial report that at least 14,000 voters participated in the pilot test is good news, it is also the case that simply having a very large sample does not make problems like selection bias disappear --- it can make it easier to identify and control for, but it cannot be ignored even in a large-sample study.
2. **Determine the extent of usability problems**. The two most pressing problems in my analysis are the issues associated with the use of smart cards as well as the

general failures associated with the optical scan voting system ("LOP"). It might be necessary to conduct further usability tests to resolve the open questions about why these issues appeared to arise with such regularity during the pilot testing.

3. **Take a careful look at the optical scanning system**. However, I do not think that because significant issues arose during the testing of the optical scan voting technology that it should be removed from consideration at this point. While analysis of the evaluation data might lead to a different conclusion, at this point it would be worthwhile to work to improve the "LOP" voting system, if the problems can be isolated, and have an improved optical scan system undergo some form of future (perhaps more limited) additional testing.
4. **Develop and implement procedures for physical security**. Before additional system development and testing is initiated, I recommend that additional security procedures be developed and implemented to increase the physical security during system revision and development. These would include, but not be limited to, stronger office security, removing critical development systems from local- or wide-area networks, and developing systems for monitoring, logging, and supervising assess to locations where system revision and development occurs. It is also important for me to stress that the project development team's emphasis on open-source software is commendable, and it should continue to be an underlying feature of future system development.
5. **Develop and implement procedures for pilot testing security**. In future pilot tests, methods and procedures for poll site security should be developed and implemented. These should include documentation of all contact that poll workers have with voting technology, external logging of all manipulations of voting technology (during set-up, balloting, and end-of-election activities), and supervision of all manipulations of the voting technology by multiple poll workers or observers.
6. **Continue pilot testing, but narrow down the number of options being tested**. As I noted at the beginning of this report, the pilot testing conducted by the City of Buenos Aires on October 23, 2005 was highly ambitious and well designed. It involved the testing of four very different voting systems, in a planned 53 sites throughout the city. The downside of such an ambitious pilot test is that there are many variables here that differ, across the four very different voting devices, across the poll sites, and even perhaps over the course of election day. It will be difficult to control for all of these variables in an evaluation, not matter what the size of the participant sample, especially when there are other issues like self-selection to deal with. Hopefully the just-conducted pilot test will help the development team focus on a smaller range of voting system variations for future testing, so that these future tests can be more tightly controlled and thus that treatment effects (variation across voting systems) can be more easily identified.