

Testing Alternative Hardware and Ballot Forms

Michael W. Traugott
Center for Political Studies
The University of Michigan

Prepared for the meeting of the Working Group on Voting Technologies and Balloting,
held at the University of Maryland, March 15-16, 2002.

The focus of this paper is on possible designs for evaluating alternative hardware and ballot designs that could be used in contemporary elections. This work can be accomplished through a number of different approaches. The first would be a series of basic research studies involving a series of experimental designs in which subjects would be exposed to alternative voting devices and/or ballots and asked to evaluate them individually and in comparison to each other. This work could be complemented with focus groups, both as an antecedent to the experiments and as a follow-up discussion after systematic measures have been taken. A second possible activity would involve the development of a facility to evaluate proposed ballot-device pairings for election administration officials, based upon what has been learned from the experiments described here as well as from Parts 1 through 3 of the project. This would be an applied research function designed to provide practical assistance under real time conditions to people who are responsible for administering elections.

In the spirit of a discussion paper, I raise a number and variety of issues here with an indication of their potential research value. They do not all have an equal cost or feasibility. However, they represent a range of possibilities for empirical data collection that would inform decisions about evaluating alternative voting devices and inform a decision about whether and how to proceed to the stage of providing direct assistance to election administration officials.

The Design of the Experiments

The central purpose of the experiments would be the evaluation of alternative voting devices with regard to a series of outcome measures or dependent variables. The number of such outcome measures would be relatively limited, and the focus would be on the internal validity of such designs, that is the establishment of high levels of confidence in the existence of causal relationships. Through the selection of study sites, some attention would be devoted to the types of subjects employed in the experiments and their reflection of the population as a whole, but external validity would not be a significant component of these designs. Recruitment site selection could be roughly related to socioeconomic status and thereby likely technical competence or sophistication and literacy levels. More elaborate attempts to reflect the characteristics of the general population would have to be accomplished through other means and data sources, including information derived from the naturally occurring experiments that arise from different jurisdictions that employ different kinds of devices and ballot forms.

In general, the experiments would be based upon post-test only designs. That is to say, we would not employ pre-tests that might involve such questions as who the subject intended to vote for because of the possibility that they might sensitize them to the purpose of the experiments and create demand characteristics that would in fact alter their behavior in the “voting booth.” We would have to construct or devise something analogous to a control group based upon a simple ballot design and device such as a paper ballot.¹

¹ One possibility would be to construct a Voter’s Pamphlet for the (fictitious) candidates and issues on the ballot and permit the subjects to study this before making their choices. They could even take a marked

My own feeling is that the issue of ballot forms will have to be treated somewhat separately and distinctly, folded in to the experiments at a second or third stage. That is to say, the number of alternative devices is relatively small, and key elements identified in Herrnson, Niemi, and Richman, as well as some identified below, provide a useful starting point around which to design the assessments. The addition of multiple ballot forms will add a significant degree of complexity to the designs that will vastly complicate the experimental designs, in terms of the number of cells and therefore the number of subjects required.

As a result, I propose that we begin the experiments with a somewhat simple, standard ballot because this will eliminate the potential for a complicated factorial design. Its properties should include multiple offices and issues, incorporating the possibility for roll off. It could involve a high salience office like President, a major statewide office like governor or Senator, district-level legislative offices with partisan candidates, district-level nonpartisan offices such as judgeships, and referenda.² There should be an option for casting a straight-ticket vote in the partisan part of the ballot. Eventually, we would want to move toward more “realistic” ballots, even prototypes that might or were used in some jurisdictions. The results from experiments with those ballots could be compared against the “baseline” of the standard or prototype ballot we would start with.

form with their choices into the “voting booth.” If the ballot is not too complex, a paper ballot marked with X’s could be used as the control.

² We will always have problems in using devices that people are unfamiliar with, but we should also be sensitive to how we conduct ballot design experiments using offices or specific candidates that the subjects are unfamiliar with as well.

The Main Measures of Suitability

One of our main tasks will be to develop a set of “standard” measures by which to evaluate the “suitability” of different devices. Two of the main measures that are relatively obvious will be *accuracy* (recording of preferences or votes as the voter intended) and *efficiency* (the amount of time required to cast a vote). Of equal importance, we would want to measure the individual voter’s assessment of the ease of use of the device and confidence that it recorded the vote accurately. And we would also like to measure less obtrusive indicators like the level of straight-ticket (or split-ticket) voting and roll off across the ballot.

There may be other measures that are not so obvious that could be related to voting devices’ ease of use and voter acceptability. Herrnson, Niemi, and Richman have a proposed list of such “usability” criteria on pages 2-3 of their paper. In order to develop a sense of the relative value of these criteria and what others might be, we should begin the project with a series of focus groups directed at two topics: how the participants view the voting process generally and what expectations they have from it, as well as their evaluations of specific devices that they could view and use (live or through video excerpts as a control) during the course of their group meeting.

Whatever the final set of measures turns out to be, they should ultimately be taken in an experimental setting unobtrusively as well as through direct questioning, allowing for appropriate IRB approval. That is to say, it will be useful to employ video to record the subject’s interaction with the devices and ballots, as well as to use post-voting questions.

One of the more illuminating photographs in prior work, admittedly based upon an unusual if not extreme example, was of shorter women who did not vote on referenda in a particular type of booth because the issues were at the top of the ballot in a very high device.

The Experimental Settings

There is a tension in designing the experimental setting for this work. On the one hand, we need a central location where a significant number of voting devices can be assembled in a single location. This will facilitate designs where combinations of disparate devices can readily be assembled. At the same time, the accumulation of these devices in a central location limits the possibilities for recruiting subjects. As a result, we should consider the possibility of two different kinds of experimental settings. First of all, we should have one or two fixed location sites that, while dispersed, offer the possibility of assembling large amounts of equipment.

At the same time, we should consider the development of one or more mobile units that could be sent to malls and shopping centers in a variety of locations. These would essentially be “mini labs” or smaller versions of the main experimental setting, perhaps set up inside a standard trailer truck the way that some manufacturers display their wares on a traveling basis. One advantage of such a mobile unit would be the ability to move the voting devices around to different locations in order to recruit subjects with more diverse backgrounds, especially by socioeconomic status.

The main purpose of these facilities will be to test a series of hypotheses about which devices “outperform” others or have higher scores on the criterion measures developed in the first stage. The hypotheses will be developed from the initial focus group conversations and be informed by the literature on human-computer engineering. It is unclear whether these experiments should be conducted on a single device at a time or they should be designed on the basis of a series of explicit comparisons.

The number of subjects required for the experiments will be relatively large, primarily as a function of the number of devices to be tested. A typical cell size would require 40 to 50 subjects to develop stable estimates of the dependent variables. The number of cells would depend upon the number of devices, paired comparison of devices, and/or device-ballot combinations. Ultimately, the number of subjects could have a significant impact on the total project budget under the assumption that we would pay the subjects an incentive or participation in the focus groups and experiments.

A Ballot Evaluation Program

The main problem with the “butterfly ballot” experience in Palm Beach County was that a well-intentioned election official designed a new ballot form with the goal of assisting a specific group of voters. The manipulation of one important factor, font size, was accomplished without enough attention to its consequences for another important factor, the alignment of names and columns. A simple pre-test of the new design might have quickly highlighted the potential confusion that the new ballot design could create. The purpose of the second part of the project would be to develop facilities to do such

pretesting in a rapid turnaround basis that would provide election administration officials with feedback to assist them in performing their functions.

Based upon the information developed from the experiments and the work with human interfaces with technology and ballot design, we should strive for the development of a set of ballot-device design principles. If that is in fact possible, the next step would be to codify these principles. Ultimately this would be in the form of software, but in the shorter term it might involve a facility in which the ballots could be tested on actual devices. The goal would be to develop a “virtual laboratory” for evaluating new ballot designs as they might be used with specific voting devices. This stage of the project should be developed in conjunction with an advisory group of election administration officials.

Summary

We can develop a research program to test the effects of voting devices and ballot forms on voters’ behaviors and attitudes. This program would begin by focusing on the testing of hypotheses to support basic knowledge accumulation, but it clearly has a number of applied consequences that we should move toward at a later stage of our work. Because of inherent design complexities, I propose that we focus primarily on devices rather than ballots, although ballots will have to be introduced at a second or later stage of the project, aided by an advisory group composed of election administrators.

The best way to pursue the research at the first stage of the project is through carefully designed experiments. They should be organized and defined by preliminary focus groups. And a laboratory setting can be established in one or two fixed locations and supplemented by a rolling facility. In some cases, it should be possible to conclude experiments with group conversations that approximate focus groups.

These design elements will provide us with strong internal validity but may leave something to be desired with regard to external validity. Therefore, we may also require additional data collection activities in the form of ecological analyses of naturally occurring experiments as well.

A distinguishing feature of this research program is that we can develop a sound empirical basis for establishing a kind of consulting service for local election administration officials. Given the knowledge base developed in the first stage, we should be able to move ahead to a system for providing rapid response answers to administrators looking for information about the suitability of equipment adoptions and, eventually, changes in their ballot designs.