INDEPENDENT EVALUATION OF CAM/C DEMOCRACY SURVEYS

William Mishler, Ph.D.
Michael Bratton, Ph.D.

USAID FROM THE AMERICAN PEOPLE

MSI MANAGEMENT SYSTEMS INTERNATIONAL

William Mishler is Professor of Political Science and former Department Head at the University of Arizona. He is Associate Director of the New Europe Barometer, a cross-national survey of 13 former communist regimes in Central and Eastern Europe and also serves as a consultant to the USAID-funded, *Knowledge, Attitudes and Practices Surveys* (I & II) in Sri Lanka. A former Director of the Political Science Program at the National Science Foundation, his most recent book is *Democracy and its Alternatives: Understanding Post-Communist Societies* (1998) with Johns Hopkins University Press.

Michael Bratton is Professor of Political Science and African Studies at Michigan State University. His most recent books are *Public Opinion, Democracy and Market Reform in Africa* (2004) and *Democratic Experiments in Africa* (1997), both with Cambridge University Press. He co-directs the Afrobarometer, a cross-national survey research project in 18 African countries and is a member of the Executive Board of the Global Barometer, a collaborative cross-national survey project including more than 250 surveys collected across 55 countries in Africa, Asia, Latin America and post-communist Europe.

Table of Contents

This report provides an analysis and evaluation of the Democracy Surveys in Central America and Mexico, plus Colombia (hereafter CAM/C) conducted by the Latin American Public Opinion Project (LAPOP) founded and directed by Mitchell A. Seligson, Department of Political Science, Vanderbilt University. The CAM/C surveys were conducted with the assistance and support of the United States Agency for International Development (USAID), which commissioned them as part of a coordinated strategy both to promote greater democracy and prosperity and to bolster confidence in democratic institutions among citizens in the region. The surveys were conducted by LAPOP, a multi-country, multi-investigator collaboration using state-of-the-art survey research methods to investigate the nature and extent of democratic attitudes, values and behaviors in Central and South America. The CAM/C project is intended to supply high quality data on public opinion in the region that can be of value to citizens, opinion leaders, public officials, NGOs, academics, and the international donor community, among others with interests in advancing and protecting democratic institutions and practices.

Overview of the Project

Our overall assessment is that the CAM/C Democracy Surveys were very well planned, quickly and effectively implemented, and appropriately interpreted. Indeed, LAPOP as a whole and the CAM/C project in particular appear to be meeting LAPOP's ambitious goal to "produce quality survey data (and analysis) that matches the highest standards of academic research in the US and Europe." An important part of this quest is to "make our methodology transparent and replicable" (p.x). CAM/C Project methods are meticulously documented and conform to the highest international standards in virtually all respects. Sampling frames are carefully designed and executed to insure that all countries have a national probability sample thereby allowing valid national and cross-national comparisons to be made. Data have been painstakingly coded, recorded, and double-checked for accuracy. The data have been rigorously analyzed and cautiously interpreted. And the conclusions drawn are well supported by the evidence presented and provide valuable insights into the problems and prospects for sustaining and extending democracy in the region.

Most of these activities occurred during a narrow window between the planning meeting held in Panama in January 2004 and the plenary session held in Costa Rica in June 2004 to review preliminary drafts of the individual country reports. The various stages, including data collection, data entry and data analysis, and report writing, were conducted virtually simultaneously in all countries during the intervening months. By doing so, the CAM/C project achieved the synchronicity that is greatly preferred in cross-national surveys (especially for data collection) in that it helps to control for potentially confounding factors related to time, which can occur as a result of sudden, unexpected events that occur in the international environment. In this regard, the CAM/C project follows the 'best practices' of the Eurobarometer, New Europe Barometer and Latinobarometer - large cross-national survey programs, all of which conduct fieldwork simultaneously in their respective regions. It is superior in this regard to the current practices of the Afrobarometer and Asia Barometer whose fieldwork, to date, has necessarily been spread over longer periods due to difficult fieldwork logistics and the gradual and uneven disbursement of funds from multiple donors. While adhering to a tight schedule (and despite a transfer of project's home base from the University of Pittsburgh to Vanderbilt University), the CAM/C survey appears to have been effectively managed and professionally conducted.

CAM/C project documentation is admirably candid in drawing attention to the oft-obscured fact that all survey data contains error from multiple sources, including not only sampling error but also

measurement and data management error. Full disclosure is also provided about non-coverage of inaccessible geographic areas or language groups and about the procedures for imputing missing values in the data set. The CAM/C documentation does not point out, but we should, that sampling, measurement and data management error is inherent in all research -- in the "hard" sciences as much as in the social sciences. Such error is not inherently any worse in survey research than in other observation and data collection methods. Indeed, the National Academy of Science officially endorses the use of probability samples and survey research as a more accurate method for gathering data on U.S. national population characteristics than traditional census procedures. This is not only because the 'sample' that results from the hundreds of thousands of people missed by the census is non-random and fraught with systematic bias, but also because the size of the census enterprise virtually guarantees more data management error. Importantly, the CAM/C project is rigorous in confronting and minimizing potential sources of error. The project provides the standard disclaimer about confidence intervals (plus or minus 2.5 percent on national samples of 1500), but also explains how even the best probability samples are likely to be wide of the mark at least some small proportion of the time (in this instance, 5 times out of a 100.

In our judgment, the CAM/C project also sets high standards for data management. In each country, all data were entered twice and the files verified against each other. An audit, undertaken at a central location, randomly selected 100 original questionnaires from each country to be physically checked against the electronic data files. Documentation provided by Project Director, Professor Seligson, indicated that survey firms in all countries were contractually obligated to reenter all data if tests of a 10% random sample found as few as 1 data entry errors per thousand keystrokes (about 1000 keystrokes were required per questionnaire). In most countries the error rate was found to be less than 1 in 10,000 keystrokes. Problems were encountered in two countries, Mexico and Guatemala, both of which re-entered and re-checked all data, de novo. After this was done, samples from both countries confirmed that both data sets surpassed the project's quality standards. We know of no other international survey project that maintains such high standards for data entry. Data entry errors of less than 1 in 10,000 are statistically trivial. They also almost certainly are random which insures that they will have no effect on the statistical results given the large samples sizes used in the CAM/C surveys.

We concur emphatically with the LAPOP and CAM/C project commitment that, "publicly funded data bases should be available to the public" in a timely fashion. At the time of this writing, however, we have been unable to locate the CAM/C databases on the USAID web site or elsewhere in the public domain. We assume that the data will be made publicly available in the near future and urge that those responsible for the project do so quickly. It also is critical, in our opinion, that the national and cross-national reports produced from this project be disseminated to as wide of an audience and as quickly as possible. While some aspects of public opinion are highly stable, others are dynamic and capable of changing rapidly in response to changes in the national, regional and international environments. All data have a limited shelf life; to maximize their value and their impact, these reports must be disseminated immediately.

For the moment, the public face of the CAM/C Democracy Survey consists of the eight country reports and one cross-national summary. At USAID's insistence, these reports use simple statistics to interrogate the data. We concur that such reports are much more readable when statistical results are described in the text with supporting bar charts and bivariate tables. In non-academic publications, it is entirely appropriate to consign more advanced, multivariate statistics (such as regression tables) to footnotes or appendices that can be consulted by interested technical specialists. At the same time, we applaud the CAM/C project participants for their diligence in including the multivariate results. These bolster the confidence of academic specialists in the project's analyses and conclusions and insure that

the bivariate results are not simply statistical artifacts of the failure to control for potential confounding influences.

The results reported by the CAM/C project have a high degree of face validity. They are generally consistent with those reported in other, narrower (in both substantive and geographical scope) surveys in the region and also conform broadly to the results reported in studies using survey data from the Afrobarometer, Asia Barometer and New Europe Barometer, among others. A study whose results largely contradicted the knowledge generated in previous research simply would not be credible. At the same time, however, the CAM/C project goes well beyond existing studies both in Latin America and elsewhere and provides important new insights into the nature and extent of democratic attitudes and values in the region. The project's innovative measurement of democratic support is a case in point (as elaborated later in this report), as is the careful measurement of corruption, and the unusual but welcomed focus on public attitudes toward local government and politics.

All large survey projects confront difficult tradeoffs between scientific ideals and practical realities, between project breadth and depth, between replication and innovation. Inevitably these choices provide ample opportunities for second-guessing, especially after the fact when the results of the choices are known. It is inevitable that a project as ambitious as this would be subject to second-guessing and, indeed, there already has been some limited criticism about some of the initial results. As relatively objective outsiders to this project and region, however, we are impressed not only by the high standards followed by the CAM/C project at all stages, but also by how little second-guessing these provocative results have generated. Our overall assessment is that this is a well conceived, executed and analyzed set of surveys that ought to provide those interested in democratic governance and citizenship in Latin American and beyond a rich source of reliable information that is currently unmatched elsewhere in the region.

Specific Issues

The terms of reference for this evaluation asked that we give special consideration to the following issues: Do the sample designs adhere to scientifically accepted norms in the field of political science? Do the questionnaires tap into the kinds of things political scientists and democracy researchers want to know about the attitudes and behaviors of the mass public? Do the question items themselves conform to best practices in the field of survey research? Do the scales used conform to good/standard practice? Is it standard practice to analyze survey items and scales like the ones used in this survey with parametric statistics, or are non-parametric statistics to be used instead? And, since the interviews averaged 42 minutes each, is the time required for respondents to complete the survey in accordance with standards for the social sciences for face-to-face, in-home interviews? Finally, where USAID's results do not completely conform to Latinobarometer findings, whose data are more reliable and credible?

1. Sampling

The construction of a valid national probability sample is the first and, arguably, the most important step in conducting survey research. Without it, nothing meaningful can be inferred about the population based on the sample results. The basic idea of a probability sample is simple; a good sample is one in which every individual in the target population has an equal chance of being selected to be interviewed as part of the survey. The obvious way to construct such a sample is by randomly selecting individuals from the population. But this simple idea is difficult to achieve in practice, especially where that population is culturally diverse or large and geographically dispersed as in most Latin American countries. In such contexts simple random sampling usually is impossible because of the absence of an up-to-date nation-wide list of citizens and their addresses and also because the cost of administering such a survey would be prohibitive.

To compensate, most high quality surveys rely on stratified, cluster sampling. Typically, such samples randomly select a certain number of large units (states, counties, electoral districts, etc.). Within this first geographic stratum a second random sample is selected from amongst geographic subunits (city blocks, rural postal routes). Often as many as four or five strata are sampled down to the level of the neighborhood, which is typically the Primary Sampling Unit or PSU. Next, individual households within neighborhoods are randomly selected. Finally, a randomly generated "Kish Grid" typically is used to insure the random selection of the member of the household to interview. Multiple attempts typically are made to interview the member of the household identified by the Kish Grid. If after repeated efforts to schedule an interview the selected household member is unavailable, then a replacement household is selected and a new Kish Grid used to select the member of the new household who is to be interviewed.

While stratified national probability samples are the international standard, their use in Latin America remains uneven. In many countries in Latin America, the cost of conducting surveys in outlying rural areas has led some firms to limit samples to urban areas. Quota samples also sometimes have been used, according to which different population groups are interviewed in proportion to their population size, but the selection of people for inclusion in the quota is non-random. Regarding the Latinobarometer, in particular, its sampling methods traditionally have included a combination of stratified probability methods, and other less expensive and less effective methods depending on the country and year. We have been told, however, that the Latinobarometer has improved and standardized its sampling methods over time and has extended its coverage into rural areas in certain countries. But we do not have access to the same detailed information on Latinobarometer methods as we do for CAM/C. Thus, while we are certain of the quality of the CAM/C procedures, we suspend judgment about Latinobarometer samples.

Based on the information provided, it appears that the CAM/C surveys closely adhere to international 'best practices' with regard to sample selection. All of the national surveys use similar stratified cluster sampling procedures, distinguishing, first, between rural and urban areas and then using standard (to each country) census units and sub-units, selecting a random sample of each, in turn, down to the level of neighborhoods, at which stage a sample of households is identified. At each stage the selection of units follows the international standard of "probability proportionate to size" which ensures that citizens in different types of PSUs have an equal probability of being included in the survey. The one deviation from standard sampling practices occurs at the household level where the use of Kish Grids is

discarded in favor of a quota procedure. CAM/C sampling procedures divide the households into four groups: Males 18-30, Males 30+, Females who go to school or work outside the home, and Females who are housewives or retired. Quotas are established for each group as determined by census data. The rationale for using quotas at the household level is straightforward and reasonable. The difficulty of finding a specific household member at home and the need to make numerous call backs to interview a randomly selected individual greatly adds to the cost of the survey and inflates the non-response rate. This is especially true in rural areas that are hard to reach and where public transportation and accommodations are limited. The effects of quota sampling at the household level on the probability characteristics of the sample are likely to be very small (unlike when quota procedures are used for level sampling strata). Nevertheless, quota sampling, even at the household level, should be a last resort. While the rational for quota samples at the household level in remote rural areas is compelling, consideration should be given in the future to using Kish grids or other random methods to select household respondents in urban and in more accessible rural areas.

Nevertheless, we commend the CAM/C team for openly engaging in the debate about other "design effects." When descriptive statistics are presented from survey data (e.g. totals, averages, percentages, differences) researchers usually report only a standard margin of sampling error (in this case, plus or minus 2.5 percent). But this conventional confidence interval assumes an unrestricted random sample. It does not take into account the fact that the primary sampling units in national samples have been stratified and clustered for practical reasons like survey cost and fieldwork logistics. Nor does it take into account that the process of weighting pooled data (see below) is itself a form of stratification. These design effects, which can be readily calculated, invariably widen the confidence interval and increase the need for caution in interpreting significance tests. So, when the Costa Rica team reports design effect statistics, they are establishing a standard to which all survey researchers should aspire.

Of course, sampling error of +/- 2.5% means, on average, that 5 out of every 100 point estimates based on the CAM/C surveys are "wrong" in the sense that their "true" values lie outside the range. Most of these will miss the mark by a very small margin; an estimate of 43% may actually be 46% (perhaps 39%), which statistically represents error but substantively is very close to the estimate. Proportionally many fewer estimates will miss by larger margins in ways that might affect substantive conclusions. It is possible to reduce the error rate by increasing the sample size, although at great additional expense for small improvements. It is also possible to reduce the error rate by conducting repeated surveys of the same populations over time and, then, either 'boot-strapping' the resulting samples or using Bayesian statistics. On balance, the error rate of the CAM/C surveys is well within scientific norms, is comparable to or better than what could be achieved by other methods and can be reduced, still further, by fielding future surveys in these same countries at regular intervals.

When pooling national data sets in order to make cross-national comparisons as the CAM/C project is intended to do, a question frequently confronted is whether to weight country samples according to relative population size (so that there are proportionately more citizens in the sample from larger countries) or to weight all countries as having equal populations. There is no single, universally "right answer" to this question. Proportional weighting is usually more appropriate where the interest is in drawing inferences about the entire population of citizens (e.g. all Latin Americans or all CAM/C citizens). Equal weighting, however, is more appropriate where the primary interest is comparing the attitudes of citizens across countries. Given the concern with comparing democratic attitudes, values and behavior across countries, CAM/C appropriately weights country samples equally.

2. Questionnaire Construction and Question Items:

In our judgment, the core questionnaire for the CAM/C surveys is a well-designed research instrument that measures public attitudes to democracy and related behaviors in a manner both theoretically relevant and contextually useful. As such, it manages to address the needs of scholars and practitioners alike. The instrument as a whole strikes a good balance among several potentially incompatible goals: broad topical coverage, deeper conceptual probes, and an interview of manageable length. In short, it represents an efficient set of compromises that together reflect the designers' considerable collective experience at planning, administering, and analyzing democracy surveys in the Latin America region.

The questionnaire covers an appropriate scope of topics for democracy research. Main themes from the democratization literature are well represented, including popular attitudes to democracy, state legitimacy, political tolerance, associational life, social and institutional trust, the rule of law, evaluations of local government, experiences of crime and corruption, and political participation including voting behavior. Within a questionnaire of this length, it would be difficult to cover other thematic territory without sacrificing one or more of these topics. With respect to subject matter, we think that the questionnaire designers have established just about the right set of priorities. With slight variations, the issues listed above are the ones that tend to preoccupy researchers elsewhere in the world in democracy surveys.

One can always quibble about the details of a questionnaire. Even lay people often assume that they know better than professionals when it comes to survey question construction. USAID and its researchers are therefore free to accept or ignore the following minor queries:

- What are the criteria for "urban" and "rural" areas? Are they the same in each country?
- Does the filter for those who answer "No" on Q. CP5 lose data on respondents who might recall some involvement if they hear Qs CP5A-5E?
- Would it be worth trying to capture more information on Qs 1T2-3 by asking "how often?" do people care/take advantage and by using a 4-point response scale: e.g. never, occasionally, often, always?
- The sequence of questions JC1-13, which are predicated on the idea of a "justified" military coup, read to us like leading questions.
- How do you handle illiterate and/or innumerate people who do not understand the 7- and 10-point scales on the prompt cards? It is not enough to instruct the interviewer to "make sure the person understands correctly" (p.7).
- Since Q.B47 is part of a battery on trust institutions, shouldn't it refer to the Electoral Commission? A question on the honesty of elections ("free and fair?") would then belong in the battery on the respondent's last voting experience.
- D1-4. Is the reference to anti-democratic social elements clearly understood? Or will respondents simply think about legitimate opposition parties and voices?
- Despite what is said about referring only to "bribery" in the overall project report, the term "inappropriate payment" still remains in EXC13.

- Are the ideological terms "left" and "right" part of common parlance in the CAM/C region? Are they, unlike "liberal" in the US and "conservative" in the former Soviet Union, stable concepts?
- The questions on electoral and policy reform presuppose a lot of knowledge. Are you confident that code 88 is really capturing all the "don't knows"?
- Q10. Income data in this format is notoriously unreliable. Since you use material possessions as a proxy for wealth in your reporting, why not drop the income measure?

Within the CAM/C research network, how participatory was the process of questionnaire construction? The study Director, national research teams and USAID democracy officers engaged in an extensive planning process for these surveys, which included detailed negotiations on questionnaire design. In the end, agreement was reached on a common core of questions that would be identical across all of the countries. A series of optional modules also were constructed in order to create common measures for issues of concern in some countries but not others. And individual country teams also were allowed to add a battery of questions unique to their individual country if they desired to do so. Although time consuming, broad consultation on survey content is an essential step, not only to build ownership of the project and its results, but also to customize the instrument to fit local contexts and user needs. It would have been useful to have more information on the ease (or difficulty) of coordinating the different indigenous versions of the questionnaire and to know why some questions were apparently not asked in some countries. Was this simply a function of belatedly adding countries that had been surveyed earlier (e.g. Bolivia 2002)? Or was it ever a function of slippage within the research network in the process of arriving at a final, agreed-upon instrument? Despite these general concerns, our overall observation, based on a review of the planning workshop agenda and communications about finalizing the questionnaire, is that the project seemed to achieve a balance between two essential requirements: local partner participation and central management control.

In addition to collaborating closely with the national teams on questionnaire construction, the study Director invested heavily in the development of human capital, devoting considerable time and effort to training members of the national teams in the finer points of sampling theory and practice, data management and presentation, and statistical methods. These investments should continue paying rich dividends in the form of higher quality survey research and analysis in the region long after the CAM/C project is finished and its results fully disseminated.

3.   Measurement and Scaling:

All measurement is imperfect. While an automobile speedometer provides a reasonably good measure of speed, its readings are subject to error depending on such factors as the size of the car's tires and the accuracy with which the speedometer initially was calibrated. In social surveys, where the objects of measurement are more subtle and involve concepts as support for democracy and political trust, obtaining valid and reliable measures is even more difficult.

One important way to confront the problem of measurement error is the use of multiple measures. Rather than asking a single question about whether people support democracy and accepting the responses as gospel, there are significant benefits to asking multiple questions about support for different aspects of democracy and using the responses to "triangulate" on the respondent's "true"

attitude. Responses to survey questions always include a mixture of valid information or "signal" about an attitude and "noise" which includes both irrelevant information and error. Different questions about the same attitude generate different degrees of signal and different types and levels of noise. Scaling is a set of procedures whereby multiple measures of a concept are combined statistically into a composite measure that manifests a better signal to noise ratio than any of the component measures individually.

There are different ways to construct scales, but all of them require that the components of the scale be substantially related both theoretically and empirically. Assessing the theoretical congruence among multiple measures is essentially a judgment call whose validity depends substantially on the investigators' bona fides in their subject area. Whether multiple measures are empirically linked, by contrast, is a matter of statistical analysis.

The CAM/C project is meticulous in documenting the procedures used in scaling key concepts. The scales developed are consistent with prevailing theory, use appropriate statistical procedures, and adhere to professional measurement standards. The CAM/C procedures in this regard stand in contrast to many commercial polls in Latin America and elsewhere which often rely on single questions to measure complex concepts or which aggregate multiple measures without due regard to their theoretical or empirical coherence or 'fit.'

Regarding the statistical analysis of the resulting scales, the CAM/C reports are quite conservative. Where included in the reports, the multivariate analyses appear to follow standard practices in the discipline. The principle technique employed is OLS regression, which remains the statistical workhorse of the profession, especially in the analysis of survey-based data. These analyses meet, and frequently exceed, contemporary professional standards. For example, the use of 'robust standard errors,' while commonplace in research on American public opinion, is just beginning to be applied in the better survey work done outside the United States.

Regression analysis and Logit/Probit, the other multivariate procedure used widely in these analyses are both parametric statistics. Indeed, non-parametric statistics have not been widely used in survey research since the middle 1960s. The reasons for this are simple. Given surveys with large numbers of respondents (500+), that employ at least ordinal level measures, include a substantial number of response categories (typically five or more), and have a reasonable distribution of responses across response categories, the variables produced closely approximate a normal distribution which allows for the use of parametric statistics. Since parametric statistics are much more powerful than non-parametric statistics and provide much greater and more easily interpreted information, they almost always are preferred to non-parametric statistics. While it is possible that some ultra conservative mathematical statisticians (though certainly not all) might object to this, the use of parametric statistics in the analysis of survey data is clearly the U.S. and international survey standard.

Finally, regarding the "risk of error" in multivariate analyses, it is important to note that the coefficients (or estimates) of the multiple predictor variables in multivariate analyses are statistically independent. This means that the confidence interval is the same (+/- 2.5%) for each estimate no matter how many other predictors are included in the analysis. Indeed, if multicollinearity is not present in the analysis (and the CAM/C report confirms that it is not), then adding predictor variables to the analysis actually narrows the confidence intervals around the estimates, thereby increasing the likelihood that the coefficients measure the "true effects," and, typically, increasing the variance explained in the dependent variable in the process. Indeed, statistical significance tests take explicit account both of the number of cases (interviews) and the number of predictors (variables) in the analysis. Moreover, this is

the same for all parametric statistical analyses, whether of survey data, economic data, climate change data, or data on the causes of sunspot activity.

4. Survey Length:

The length of surveys is critical. Respondent fatigue can seriously degrade the quality of the data collected and greatly increase missing data. The maximum appropriate length for a survey varies with the survey context, however. The more educated and sophisticated the respondent (and the more surveys someone has answered in their lifetime) the harder it is to capture and sustain a respondent and so the shorter the survey needs to be. Similarly, the more difficult it is for the interviewer to establish rapport with the respondent, the shorter the interview must be. In the United States, where telephone interviewing in the norm and citizens are deluged with phone surveys of all sorts, the accepted norm is that a 20-minute survey is about the maximum that can be fielded without risking serious respondent frustration. However, the National Election Study, which is the Gold Standard for political surveys, has relied primarily on face-to-face interviews and found as recently as 2002 that 45-minute interviews are quite feasible. In contexts where respondents are less jaundiced about being interviewed and where face-to-face interviews are conducted in the respondents' homes, interview lengths of 40-60 minutes usually should not be problematic. The CAM/C average interview length of 42 minutes falls comfortably within this accepted range. In many contexts interviewers report that subjects are delighted to talk with them, once a certain minimum level of rapport is established. And interviewers frequently find it difficult to disengage from a respondent at the end of the interview. CAM/C interviews in Central and South America, where being surveyed is novel to most citizens, should not experience any significant respondent fatigue with interviews of 40 to 45 minutes.

Latinobarometer surveys and those conducted by commercial firms in Latin America typically are much shorter than 40 minutes and probably average 20 minutes or less. This typically is not because of concerns about respondent fatigue, however, but because their surveys are narrower in scope so there is less need for longer surveys. There is nothing inherently better or worse about long or short interviews provided that the length of the interview is appropriate to the context. Our experiences interviewing in Africa, Asia and Central Europe suggest that the length of the CAM/C surveys in Central and South America is reasonable and appropriate.

5. Most Important Findings:

The project makes a valuable and original contribution to the measurement of democratic stability. The stability of democracy is captured by "system support", a scale of five items that address popular trust and pride in a set of institutional procedures. Like other scholars, the CAM/C team chose to avoid asking respondents directly about the "d-word," since "democracy" is a concept overloaded with positive connotations (see Rose, Mishler, and Haerpfer, 1998). They also impute democratic stability from "political tolerance" (a scale of four items about support for the political rights of anti-

democratic social elements), as do other political scientists working on Russia and South Africa (e.g. Gibson and Gouws, 2003). It should be noted, however, that their "system support" is a somewhat idiosyncratic construct that is strongly associated with the work of Professor Seligson and his colleagues that has not yet traveled far *beyond* the CAM/C countries. We confirm, however, that it provides a rigorous and meaningful basis for comparisons of democratic stability *across* the CAM/C countries. The fact that downward trends in "system support" apparently served as an early warning signal of declines in democratic stability in Costa Rica is testament to the utility of this approach.

In order to test CAM/C against other estimates of democratic attitudes in Latin America, we compared results from equivalent items across surveys. Both CAM/C and the Latinobarometer ask standard questions about support for, and satisfaction with, democracy. Although question wording and the order of response categories varies slightly across surveys, we regard these items as being conceptually equivalent and, therefore, comparable. It is therefore disturbing to find wide discrepancies in the (raw, uncorrected) distributions of results from competing surveys in 2004. Whereas, across the 8 CAM/C countries, the CAM/C survey reports an average of 64 percent of respondents who say they "prefer democracy", the Latinobarometer reports an average of only 50 percent. UCA surveys conducted by the same Nicaraguan team that conducted the CAM/C surveys also find lower levels of system support. Similarly, for the 6 Central American countries (without Mexico and Colombia), the CAM/C surveys report far less dissatisfaction with "the way democracy works" (37 percent) than does the Latinobarometer (59 percent) or UCA.

While we cannot be certain without better information, we suspect that Latinobarometer's lower estimates are artifacts of its under-sampling of rural areas. In contrast to CAM/C's national probability samples that represent rural areas in their correct proportions, the Latinobarometer samples may have an urban bias. The Nicaraguan team suspects that the UCA surveys reflect the same bias. Skeptical urban dwellers tend to be much more critical than their trusting country cousins of the desirability and performance of political regimes, including democracies. To the extent that it focuses mainly on urban respondents, the Latinobarometer may be underestimating the extent of democratic sentiment in the subcontinent. An examination of the CAM/C survey data on levels of trust in the judicial system in Nicaragua, comparing urban versus rural samples, supports this hypothesis. It should be pointed out, however, that even if some Nicaraguan UCA samples have an urban bias, this does not mean that the same bias exists in the UCA CAM/C sample. Survey firms typically draw samples according to the precise specifications of their clients. Our understanding is that the CAM/C contract for Nicaraguan specifically required the construction and use of a national probability sampling frame. We would be surprised if the CAM/C Director did not exercise due diligence in ensuring that the contract was strictly followed in regards to sampling as in all others.

Reassuringly, both the Latinobarometer and CAM/C projects report that attitudes to democracy are distributed in very similar cross-national patterns. The proportions that "prefer democracy," for example, are highly correlated across surveys (Pearson's r = .874). In both surveys, Costa Rica leads the way and Guatemala brings up the rear. And if only country rankings are considered – which is the usual limited purpose to which such annual "box scores" or "league tables" are put – then the two survey series are even more consistent (Spearman's r = .958). If forced to choose a single survey's results on democratic support and satisfaction, however, we would rather rely on the CAM/C data as the more authoritative source.

Because additional elements in the questionnaire were constructed with an eye toward work being done by other scholars, there are multiple opportunities for comparisons with public opinion in world regions beyond Latin America. We estimate that perhaps up to fifty percent of the items in the

CAM/C questionnaire are, to a greater or lesser extent, conceptually equivalent to items asked in one or more of the following: the World Values Survey, the World Bank Corruption Surveys, and the Global Barometer Surveys (New Europe, Afro-, Asia, and Latino-). Although exact question wording and response categories may vary – and such differences should always be reported – a basis does exist for USAID to make careful comparisons of survey results across different parts of the world. Indeed, given the challenge of cultural and linguistic diversity that arises with global comparisons, we think that conceptual equivalence is a more appropriate goal for question design than strictly identical wording. Accordingly, democracy researchers can find in the CAM/C questionnaire a foundation for encouraging further convergence of survey research instruments across various world regions.

In terms of other important findings, we applaud the effort of the CAM/C team to put the important concept of corruption on a firm empirical footing. Rather than stopping at popular perceptions of corruption (which are often stoked by sensational press reports or unconfirmed rumor), the questionnaire probes the extent to which citizens actually experience requests for bribes. It specifies exactly which public officials are involved (see CAM/C items EXC1-19), and the CAM/C researchers put corruption in cultural context. They seek to establish where ordinary people draw the line between acts that are illicit but considered acceptable and those that are illegal and deserving punishment (see CAM/C items DC1-DC13). Indicative of the value of this innovation, the designers of the Afrobarometer used these state-of-the art questions as guidelines when developing their own Round 3 instrument in June 2004.

Another important innovation of the CAM/C study is the attention it devotes to local government. After all, this is the level of government that is most salient to the everyday lives of poor – especially rural – people in countries where the business of central government is often seen as remote, inaccessible, and hard to understand. Under these conditions, citizens develop their satisfactions (or their dissatisfactions) with democracy based substantially upon the performance of locally elected representatives and the delivery of services by local councils. It is especially important to study the performance of local governments in countries where governments (and USAID) are promoting policies of administrative decentralization. We wonder, however, whether attendance at council meetings fully captures the extent of popular participation in municipal government (see items NP1- NP1a); we prefer voter turnout in the last municipal elections (item GVB5). But the items on citizen contacts with and the perceived responsiveness of municipal officials and councilors are surely relevant to the formation of public opinion about day-to-day democracy.

The one significant gap in the coverage of the CAM/C study, from our perspective, concerns informal politics. Most of the questions in the instrument are directed at the formal aspects of political life, such as people's interactions with official institutions, personnel, and procedures. From our limited knowledge of the literature on Latin American politics, we are struck with the persistence of informal patron-client relations, which both circumvent the official agencies of the state and pervade them (Fox, 1990; Hagopian 1996; Grindle, 2000). While there is never enough room in questionnaires to cover everything, we would suggest making space in the future to explore alternative forms of participation such as contacting leaders outside the state (such as ethnic, religious, labor or business leaders) for help in solving problems. And the reader's understanding of both electoral behavior and corruption in the region would be enhanced by explicit attention to the pervasive problem of vote buying. The evidence from Africa is that informal activities of this sort are more widespread than formal modes of participation and that, paradoxically, informal patrons serve as intermediaries in facilitating popular access to elected representatives and other officials of the state (Bratton, Mattes and Gyimah Boadi, 2004).

Conclusion

The CAM/C democracy surveys appear from our independent vantage to have succeeded in their goal of achieving the highest standards for academic research prevailing in the United States and Europe. The design of the sampling frame, the standardized content of the questionnaire cross-nationally, the quality of data collection and coding, and the rigorous analysis and careful interpretation of the results closely follow the best practices in the field and represent a clear improvement over previous efforts to undertake multi-national surveys in the region. The strong leadership provided by the project director and the hard work and close collaboration among the research teams from each of the CAM/C countries appear responsible for this success.

The high quality of these surveys means that government officials, public policy specialists, academic researchers, NGO officials and ordinary citizens in the region can have considerable confidence in the data and in the analyses, conclusions, and policy recommendations based on them. In addition to these obvious and immediate uses, these data are valuable as well because they establish a credible baseline of public attitudes and values against which future trends in public opinion can be reliably compared. In the end, this may be the most important legacy of these data. Public opinion, after all, is rarely static but is constantly shifting in response both to long term changes in the population and in peoples' life situations and to short term influences from new political leaders, public dialogue, and unexpected social, political and economic events. As a consequence, public support for democracy is something that needs to be continuously reinforced and regularly monitored. Thus the full potential of the CAM/C surveys will only be realized in future years as additional high quality surveys of the region build on this impressive beginning.

Appendix 1. List of Documents Reviewed

Mitchell A. Seligson, *The Political Culture of Democracy in Mexico, Central America, and Colombia, 2004*

Juan Carlos Rodríguez-Raga, *The Political Culture of Democracy in Colombia, 2004*

Sample Description for Guatemala.

Sample Design Guidelines

University of Pittsburgh, Democracy Audit, Questionnaire, Version 7.3, February 9, 2004

University of Pittsburgh, Institutional Review Board, Approval

Contract Guidelines and Technical Requirements for Teams

Agenda, Final AID-CAM/C Meeting, Hotel Bristol, Panama City, Panama, January 11-14, 2004

Agenda, Mid-Term Meeting of CAM/Colombia Researchers

CVs of CAM/C Senior Scientists

SPSS Data and Chart Templates

PowerPoint help files for SPSS

Precision of Results, memo on design effects

Trust in the Judiciary, responses from Seligson

Corporacion Latinobarometero. 2003. Ficha technica, Latinobarometero, 2002.

Corporacion Latinobarometero. 2004. Informe-Resumen Latinobarometro 2004, Una Decada de Medeciones (www.latinobarometero.org)

"Democracy's Low-Level Equilibrium: The Latinobarometro Poll," *The Economist*, August 14, 2004.

United Nations Development Program, *Democracy in Latin America: Towards a Citizens' Democracy: Ideas and Contributions*, New York, UNDP, July 2004.

Article by Andres Oppenheimer (*Miami Herald,* May 13, 2004) and responses by Dante Caputo, director of the report (*Miami Herald,* May 18, 2004) and in statement entitled "Aclaración del PNUD a las observaciones efectuadas por Latinobarómetro sobre el uso de datos 2002 en el Informe *La Democracia en América Latina,"* posted on May 14, 2004, on the report's official website (http://democracia.undp.org/Default.Asp); article by Carol Graham (*Financial Times,* August 2, 2004) and response by Dante Caputo, director of the report (*Financial Times,* August 4, 2004); and statement

entitled "About the UNDP Report 2004," of April 30, 2004, posted on the Latinobarometro website (http://www.latinobarometro.org).

U. S. Agency for International Development, *The Democracy Survey Database* (www.usaid.gov)

Bratton, Michael, Robert Mattes, and E. Gyimah-Boadi. 2004. *Public Opinion, Democracy, and Market Reform in Africa.* New York: Cambridge University Press.

Gibson, James and Amanda Gouws. 2003. *Overcoming Intolerance in South Africa: Experiments in Democratic Persuasion.* New York: Cambridge University Press.

Fox, Jonathan (ed.). 1990. *The Challenge of Rural Democratization: Perspectives from Latin America and the Philippines.* London: Frank Cass.

Grindle, Merilee. 2000. *Audacious Reforms: Institutional Invention and Democracy in Latin America.* Baltimore: Johns Hopkins University Press.

Hagopian, Frances. 1996. *Traditional Politics and Regime Change in Brazil.* New York: Cambridge University Press.

Leslie Kish, *Survey Sampling.* 1995. New York: Wiley-Interscience; New Ed edition .

Rose, Richard, William Mishler, and Christian Haerpfer. 1998. *Democracy and Its Alternatives* Baltimore: Johns Hopkins University Press.

Appendix 2. Scope of Work

To evaluate:

1. The survey methodology
2. The statistical analysis
3. The quality of the results (within field and compared to other LA surveys)
4. The most important findings, specific and general

Questions to answer:

1. Do the sample designs adhere to scientifically accepted norms in the field of political science?
2. Do the questionnaires tap into the kinds of things political scientists and democracy researchers want to know about the attitudes and behaviors of the mass public?
3. Do the question items themselves conform to best practices in the field of survey research?
4. Do the scales used conform to good/standard practice? Is it standard practice to analyze survey items and scales like the ones used in this survey with parametric statistics, or are non-parametric statistics to be used instead?
5. The interviews averaged 42 minutes each. Was the time required for respondents to complete the survey in accordance with standards for the social sciences for face-to-face, in-home interviews?

Comparisons with other studies:

1. USAID's data does not always completely conform to the results in other surveys, such as the Latinobarometer. In those areas where there are discrepancies, whose data should be considered more reliable? What seems to explain the differences in results?

2. In Nicargua specifically, comment on the differential results for "confidence" in various institutions between the USAID surveys, and those recently conducted by the Latinobarometer and UCA. In your professional evaluation, which surveys are the most credible and how do they stack up against the USAID survey in terms of the methodological questions above?