**Citizen Oversight of Presidential Performance**

**Abstract**

This paper offers a new approach to measuring and evaluating citizen oversight. Findings emerge from 855 interviews with ordinary citizens (recruited between 2005 and 2012 to demographically approximate random national samples). The interviews are first used to assess the construct validity of a theory-based *matrix* designed to classify the evaluative attitudes toward presidential performance that shape citizen oversight. The validated matrix is then used to retest Brody’s 1991 finding that the results presidents get is the basis for public approval. The interviews confirm the importance of results but also show that the actions presidents take are nearly equal in importance to citizens as grounds for their presidential approval decisions. Finally, the extent of continuity and change in citizen assessments of Presidents Bush and Obama is analyzed. Evidence of both strengths and weaknesses in citizen oversight of these and other presidents is noted throughout.

**Paper presented to the Western Political Science Association, Sheraton Seattle Hotel Seattle, Washington April 17-19, 2014**

The performance of citizens as agents of presidential accountability is both important and understudied. The people are the arbiters of access to and retention of presidential power. They are the *de facto* anchor of a complex, multi-agent presidential accountability system whose members (e.g., Congress, Court, and news media, among others) are also dependent on public support (Buchanan, 2013: 33-59).  Moreover, citizen judgments underlie presidential job approval, the most-watched statistic in American politics and “a proxy for the chief executive’s power to persuade lawmakers, capacity to win re-election and ability to help or hurt in mid-term elections” (Harwood, June 14, 2010: A15). Presidents pay attention.

It therefore matters how, and how well, citizens oversee presidents.  Although there has been a significant ongoing effort to assess the general civic competence of voters (e.g., Borgida, et.al., 2009; Delli Carpini and Keeter, 1989; Campbell, et. al., 1960;) and some research attention to citizen conceptions of presidential prototypes (Kinder, et. al., 1980) and to presidential greatness (e.g., Simonton, 2008) there has been very little specific work on how Americans judge presidential performance in office. That is the focus here.

 The most important precursor to the research reported below was published more than 20 years ago by Richard Brody (1991). Brody found that news reports of presidential policy successes and failures were correlated with higher and lower levels, respectively, of Gallup Poll-tested public approval of presidential performance. Brody also found that “the clearer the results and the more news devoted to a particular topic the less likely it is that public reactions will differ along lines of partisanship” (1991: 174).  These findings convinced him that citizen oversight of presidential performance was adequate.  By “basing rewards and punishments on performance” and by “resisting partisan blandishments “(t)he American people –despite minimal attention to politics, despite a lack of information, and because of the peculiar institutions of American politics—appear to have stumbled on the key” (1991: 175). Later I present interview evidence to suggest that while citizen partisanship is on the rise, the sophistication with which citizens approach presidential assessment, though still open to criticism, is greater than Brody could discern based on his inferential evidence alone.

**Theory**

Two research questions underlie the present study.  First, can interview measures of citizen judgments of presidential performance be organized within a classification system based on historical and political science research that achieves “construct validity” (Trochim, 2008) thereby facilitating valid, theoretically grounded interpretations of interview data?

Second, do the presidential performance expectations articulated in citizen interviews emphasize “the results presidents get” as the most frequently expressed basis for judgment, as per Brody (1991)?

The two theories below express affirmative answers to these questions. Each is tested later using interview evidence.

Theory A: The Matrix

 First is an inductive “values and targets” (VT) theory of citizen judgment. The theory, derived from research literature cited below, holds that when people evaluate presidents they are applying particular values (effectiveness, morality, or prudence) to specific targets of evaluation (presidential traits, acts, and outcomes, or results). Juxtaposing these value “rows” and target “columns” creates the nine cells that comprise the “hypothesis matrix” depicted in **Table 1**. Each cell represents the unique intersection of a particular value and a particular target (a VT intercept) into which citizen judgments (evaluative attitudes) may be classified. Examples of citizen judgments that fit within each of the nine VT intercept cells are in **Figure I**. To test Theory A, coders classify the evaluative commentary of interviewees into one or another of these cells (details in methods section). If independent coders populate every cell with their classifications (however unevenly), and if they classify 75 percent or more of interviewee judgments into one or another of the nine cells, then the VT theory matrix will have demonstrated “observed pattern” construct validity (Trochim, 2008) and a theory-based system for classifying and interpreting interview testimony will have been established. Why is a percentage measure the best Indicator? Because the extent to which citizen judgments of presidential performance can be classified into the cells of Table I is most clearly revealed by percentage statistics (Buchanan, 1974; Knapp, 2009).

Matrix Origins

The “bar” is set as high as 75 percent because by the standards of explained variance in social science research, this is a relatively stringent requirement, but also because there are good reasons to expect most citizen judgments to fit the matrix. For example, the *targets* arrayed across the top of the matrix derive from presidential assessment scholarship and survey research, which suggests that the three most common objects (targets) of citizen assessments are presidential traits (e.g., Newman, 2004; 2003; Gronke and Newman, 2003; Greene, 2001) presidential actions (e.g., Harwood, September 15, 2005; March 17, 2006) and presidential performance outcomes, or results (e.g., Brody, 1991; Howell and Perry, 2004). For extensive research summaries that take some account of all three targets see Kinder (1986) and Kinder and Fiske (1986).

The origins of the *values* located on the left side of the matrix require more explanation. Note first that the three values are discernible as prominent grounds for judgments of historical presidents (Bailey, 1966) as well as modern presidents (Greenstein, 2009). Second, note that because the cells are equal in size, the matrix classification format may invite an assumption that interviewees will apply values to targets at roughly the same rate in their testimony; implying substantial if not equal influence by each value on each target (and similar classification counts in each cell). This is *not* the expectation here. Row one can be expected to attract most of the interview mentions because classified there are expectations for presidential results. Results (and the presidential traits and acts that encourage results) are indicators of presidential *effectiveness* at securing valued ends or solving important problems. It is apparent in research on presidential handling of the economy (e.g., Cohen, 1999) in pooled rankings of presidential greatness by experts (Ragsdale, 2009: 27-30) and in Brody’s 1991 finding of a strong positive correlation between presidential approval scores and media reports of presidential policy success and failure. The effectiveness row of the matrix (particularly the “action” and “results” cells) is thus expected to contain the great majority of citizens’ coded evaluative commentary.

History and research suggest, on the other hand, that m*orality* and *prudence* based citizen assessments of presidential performance predictably will occur *much less frequently* than effectiveness-based assessments. This is not because they lack importance. It is because morality and prudence concerns *tend to arise only on the rare occasions when public discomfort with a president’s chosen methods for achieving his ends is widespread and intense.* Because such events can have lasting significance, comparatively infrequent mention in years that do not feature extraordinary public reactions to presidential methods is not tantamount to insignificance. Despite their infrequent evocation by events, morality and prudence rows belong in the matrix because history shows that *either may emerge suddenly and unexpectedly as high profile performance concerns*. A classic morality example is Polk’s provocation of an unnecessary war with Mexico in order to acquire California. It sparked condemnation of Polk among most who realized the extent of his duplicity (Greenberg, 2012: 91-110; Skowronek, 1993: 155-176)*.* More recently, high profile morality and prudence concerns were raised by presidential “wars of choice” such as Korea, Vietnam and Iraq (Buchanan, 2013: 59-80). Such moral concerns are addressed *passim* by historians (e.g., Herring, 2008; Goldstein, 2008), biographers, (e.g., Caro, 2012), journalists (e.g. Savage, 2007;Rich, 2006) public opinion scholars (e.g., Mueller, 1973) and political scientists (e.g., Rudalevige, 2005)

 Moral judgments can have long lasting consequences. Prominent examples are those that followed Nixon’s dissembling about his involvement in the Watergate affair (Reeves, 2001: 609) and Lyndon Johnson’s misrepresentations surrounding the Vietnam War (Dallek, 1998: 273-277). Widespread distress at these perceived breaches of faith triggered the first steep drop in a long decline in general public trust in government, from a 1966 high above 60 percent to contemporary levels in the 20 percent range -- <http://pewrsr.ch/11mfOuT>. As for the prudence value, it drove academic and media critics and some ordinary citizens to fault President John F, Kennedy for the riskiness of his strategy for resolving the Cuban Missile Crisis, Giglio, 1991: 215; Miroff, 1976). Prudence concerns led even more elites and citizens to disparage George W. Bush’s unprovoked invasion of Iraq (e.g., Rich, 2006; One Year After, March 19, 2004: A18) for incurring what they perceived as incautious and unnecessary risks to depose Iraqi dictator Saddam Hussein.

Theory B: Brody and Results

Again, it matters how well how well citizens meet their oversight responsibility. The measure of quality used here is the extent to which citizens, by their stated expectations, encourage presidents to engage in what Donald Stokes called *valence* problem-solving. Stokes identified the perennial valence issues as prosperity, peace, and social justice: things that promote the viability of the polity (Stokes, 1966: 170-173; 1992).

Following Brody, Theory B predicts that most of the citizen expectations voiced in interviews will encourage valence problem solving. Brody’s core finding (1991:166) is represented here as Theory B:

Over the quarter century from the Kennedy Inaugural to the Reagan reelection the American people have evaluated the performance of their president on the evidence

of policy success and failure supplied to it by the press. The standards of “success”

and “failure” are derived from outcomes for which there is general agreement—

prosperity is “good,” war is “bad,” and so forth—and from the expectations set by

the president…..In the aggregate, the public seems to respond to policy outcomes,

not the means of achieving them—(the public) is pragmatic rather than ideological.

Accordingly, the prediction is that over the eight-year course of the second Bush term (2005-2008) and the first Obama term (2009-20012) the presidential performance expectations voiced by this study’s interviewees will stress social, economic and international outcomes, and not the means of achieving them

**Methods**

 Political science research on assessing presidents has rarely sought direct citizen testimony. Yet as Robert Lane (1962: 8) argues, social scientists must first understand the outlook of the people they study before they can properly interpret inferential indicators. The direct testimony used here is important for this and one other reason. Citizens who choose their own words in response to general questions give *unobtrusive indicators* of their evaluative dispositions (Krippendorff, 1980:29).

Unit of Measurement: Evaluative Attitudes

As shown in the interview protocol in the **Appendix**, respondent standards and expectations are probed extensively. Interviewees are asked to identify and explain their criteria for assessing presidents in general by first explaining how they distinguished between good and poor past presidents they identified. Then they are asked to explain their overall, as well as their economic policy-specific and war policy-specific judgments of the incumbent president on the date of the interview. The interviews measure citizen attitudes, expressed as reasons for approving or disapproving of presidential performance in these categories (1).Attitudes are evaluative dispositions—a “psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor” (Eagly and Chaiken, 1993:1. See also Krosnick, et. al. 2005). The interviews operationalize the “psychological tendencies” underlying the reasons given in response to such longstanding poll questions as Gallup’s “do you approve or disapprove of how (the current incumbent) is handling his job as president?” In the process of explaining the reasons behind their approval or disapproval of presidential performance in the three areas explored by the questionnaire, their evaluative attitudes toward presidential performance are revealed (2).

Sampling Strategy

 The interviews began in 2005. Like Richard Fenno in his small-N study of Congress members (1978: 253) the sampling procedure was to slowly build up the size of the sample and to monitor its composition. The aim was to construct a progressively larger sample, convenient and nonrandom, but also “purposive” (i.e., a sample made up of ordinary citizens systematically recruited to approximate the national distribution of four key demographic characteristics: party identification, education, age, and gender (**Table II**).Although this sampling method is not preferable to a random national sample, it is a reasonable approach for a study whose intended contribution is to promulgate a new conceptual tool (Table I’s matrix). Recruitment of demographically pre-screened participants was accomplished by undergraduate students aided by a letter of introduction from the author. 855 face-to-face interviews were conducted, recorded, transcribed, archived and coded.

Proximal Similarity

 The most common criticism of nonrandom samples is that they cannot be statistically representative of the population. True, but they can be constructed to approximate the most important characteristics of the population to which we seek to generalize, as was done here. This principle is codified in Donald Campbell’s “proximal similarity model” (Trochim, 2008). This model, conceived as a basis for estimating the external validity of nonrandom samples, argues for the *plausibility of generalizing to people in the population like those in the sample*.

The greater the degree of similarity, the greater the confidence we can place in generalizations to “similars” in the population. That is why **Table II** includes, in parentheses, demographics from recent random national samples. One poll, sponsored by the Pew Research Center for the People and the Press, is used for gender, education and age. For party identification, an average compiled from several national polls by Pollster.com is used. The comparison percentages show the proximal similarity of the present sample to random national samples on four major demographic characteristics. The most important groups are present and accounted for, which makes for a balanced assortment of ordinary Americans that serves the purposes at hand with little likelihood of distortion. Despite minor differences in group percentages, there is no reason to expect the demographic “similars” in the Pew, Pollster.com or the interview samples to be anything other than that--similar. As a check against regional bias (most interviews were conducted in the Southwestern U.S.) an on-line version of the interview was conducted in 2009 with 93 randomly selected subjects recruited within the State of New York by a research firm retained by the University of Texas at Austin Survey Research Center. There were no systematic response differences between the Northeast and Southwest regional interviews.

Longitudinal Data Collection

A new group of interviewees was recruited during each of thirteen successive undergraduate research seminars led at the author’s university between spring, 2005 and fall, 2012. The interviews span the entire second Bush term and the entire first Obama term. This facilitated identification of changes from Bush to Obama in the salience of particular evaluative criteria. **Table III** equates individual classes to “interview waves.” The table also notes that dissimilarities in the number of interviews per wave and in the spacing of waves (i.e., not every interview year features two waves) are due to the vagaries of class scheduling and student enrollments.

Interviews

Seminar students were responsible for conducting, audio recording, and transcribing 855 interviews lasting between 20 and 40 minutes each (see the **Interview Protocol** in the **Appendix).** Each seminar featured a training regimen for student interviewers that covered each of the above responsibilities as well as the need to establish a comfortable environment for the interviewee, consistency in question-posing, nonthreatening requests for explanations, clarifications and elaboration of interviewee responses to questions and the avoidance of “leading” or “steering” interview techniques.

Content Analysis

Content analysis was guided by the matrix structure sourced and explained above and depicted in Table I.

Coder Training

Coders were trained to determine which among the nine “value-target” (VT) intercepts (numbered 1-9 in Table I) particular responses to interview questions best fit*.* Examples of evaluative attitudes fitting each cell (edited from interviews) appear in **Figure I.** Training began with an introductory presentation on the origins and explanation of the matrix, followed by presentation and discussion of specific coding examples like those in Figure I. Discussion of examples for each cell continued until coders were able to identify which cell was most appropriate for classifying particular kinds of evaluative commentary. They were also coached to be alert to commentary not classifiable within the matrix (see “non-matrix classification options” below).

Training Regimen

Coders were instructed to listen to the entire recorded interview before beginning to code a particular interview. This facilitated their understanding of each interviewee’s overall approach to judgment, which increased the accuracy of classifications within particular performance categories. Coding practice sessions consisted of student coders and the instructor listening to interview tapes together, and then content-analyzing the tapes independently using the **coding sheet** in the Appendix. Thereafter, coding differences were discussed and “best classifications” identified and explained. This process was repeated until acceptable student coding competence was achieved.

Note that the coding sheet has specific locations for commentary fitting each matrix cell, within each of the four performance categories included in the interview protocol. The coding sheet instructions ask coders to “rank order the importance” to interviewees of mentioned criteria (i.e., reasons given by interviewees for their judgments of presidential performance) in each performance category. But to maximize the clarity and interpretability of the results *only top-ranked criteria are included in the findingts Tables presented below.*

Inter-Rater Reliability

 After making individual coding decisions during practice coding, participants reconvened to discuss the differences that emerged in response to particular interview tapes. This process was repeated until a high degree of convergence was achieved. At this point “official” independent coding began. Specially trained Research Assistants (members of past seminars hired for the purpose) independently coded every tenth interview recording, compared their results to those on the coding sheets completed by student coder, and calculated the extent of their coding agreement with the student coders. The extent of agreement (Inter-Rater Reliability, or IRR) was calculated as a ratio of actual to possible classification agreements between coders and RA’s. IRR averaged .90 between 2005 and 2011.

Non-Matrix Coding Options

Two non-matrix classification options are included on the coding sheet for each of the four performance categories. Criterion mentions that did not fit the matrix were assigned to either the **Partisan Agreement** category or to the “**Other”** category. Partisan Agreement, itself a value, was reserved for criterion mentions that specifiedagreement with the evaluator on partisan policy, political ideology or party affiliation as the basis for the evaluator’s approval. The noteworthy feature of this criterion is that it does not expect presidential success at valence problem-solving. It is instead a litmus test in which political loyalty is the sole criterion. That is why it is outside the matrix. The “Other” non-matrix coding option is a residual category. Classified there are criterion mentions that fit nowhere else.

These two non-matrix coding options serve two related purposes. First, they allow any evaluative commentary that was not predicted (i.e., commentary not classifiable within the nine cells of the matrix) to refute Theory A if more than 25 percent of coded commentary did not fit the matrix. The second purpose was to determine whether values additional to effectiveness, morality, prudence, and partisan agreement were influencing presidential assessment. The open-ended nature of the “Other” category allowed additional values to emerge and be counted. Any previously excluded value could demonstrate its significance by attracting a large number “mentions” across this large sample of interviewees.

In sum, coders had three reasons to take both non-matrix options seriously: 1) to be strict about excluding “non-fitting commentary” from the matrix to best assess its construct validity; 2) to accurately measure the extent of reliance on partisan agreement as a valued criterion, and 3) to allow any other evaluative themes, including additional values, to emerge and be counted

**Findings**

Concepts forged in theory that correspond to real-world observations reflect a form of construct validity called an “observed pattern” (Trochim, 2008: on line). The concept “forged in theory” here is the matrix in Table I, the values and targets hypothesis. The “real-world observations” are the interviews. The “correspondence” of the real-world observations to the concept is *the extent to which the interviews can be classified within the cells of the matrix* (as determined by the coding procedures described above). As noted, the sufficiency test used here is twofold: all nine matrix cells must be populated with coded commentary and at least 75 percent of total “number 1” criterion mentions must be classifiable in one or another of the nine cells**.**

Theory A Test

**Table IV** reports the test of the values and targets hypothesis. As shown, the “observed pattern” construct validity for the values and targets matrix is confirmed. The fact that all cells have entries increases confidence that the nine-celled structure is not misspecified. As noted, the percentage summary statistic is the most relevant indicator available of the validity of the values and targets construct as a theory based tool for classifying and interpreting citizens’ evaluative commentary about presidential performance (Knapp, 2009).. Across 855 interviews, two presidential terms and four performance categories, The VT matrix hypothesis is confirmed. The matrix accounts for 81 percent of the top-ranked evaluative attitudes expressed in the interviews. The original “threshold test” of 75 percent of variance explained is comfortably surpassed. The matrix accounts for and explains the great majority of the evaluative commentary. At 15 percent of evaluative commentary partisan agreement significantly exceeds the scant 4 percent of top-ranked commentary coded into the “Other” category, which surfaced neither values nor any other discernible “mention” patterns.

Significance

The confirmation of Theory A supports a claim that the matrix can serve as a reliable cross-time classification and diagnostic tool, suitable for tracking and evaluating citizens’ judgments of presidential performance. Once “norms” for interview response patterns can be established (the eight-year record assembled to date is the first step) it will be possible to assess important continuity and change. That requires a long-term record of citizens’ evaluative performance (ideally across generations of interviewees and multiple presidents) for comparison with evolving theories concerning how the accountability relationship between presidents and citizens ought to work. Just as presidential performance is continuously measured and judged, so too might citizen performance be continuously measured and judged, with a focus in both cases on where improvement is needed.

Theory B Test

 A finding within the validated matrix in Table IV corroborates Theory B. Specifically, the mention count in matrix cell 3 is consistent with Brody’s (1991) inferential finding concerning the importance of results. “Mentions” summed cross four performance categories show that at this aggregate level, “results” is the criterion most frequently mentioned by interviewees as the primary basis for judging presidential performance. “Results” narrowly edges out matrix cell 2, “action” as the most frequently mentioned top priority evaluative criterion in the matrix. Though small (just 3%) the mention percentage difference between cells 2 and 3 is statistically significant (p<.05). This affirms, with direct testimonial evidence, Brody’s 1991 inferential conclusion concerning the impact of policy outcomes on presidential approval. The arrival at “results,” a synonym for Brody’s “policy outcomes,” via such a dissimilar measurement approach signals convergent validity.

Action and Results

But a qualification to this corroboration of Brody is necessary. Because the matrix specifiesmultiple evaluative targets (traits, acts and outcomes) it can show that presidential *actions are mentioned by citizens nearly as often as results* as the principal basis for judgment (compare cells 2 and 3, Table IV). Table IV also shows that “action” and “results,” at more than 1000 mentions each, are far and away the most mentioned grounds for judgment in the matrix. “Action” is similar in aggregate mention frequency to “results” as a basis for judgment in every year of this eight year interview study. Brody’s study does not address presidential action. He dismisses its importance when he states that “…the public seems to respond to policy outcomes, *not the means of achieving them*…” (1991:166. Emphasis added). By “means of achieving” policy outcomes Brody means policy proposals intended to achieve outcomes. He does not mention other forms of presidential action, such as speechmaking, image cultivation, position-taking, or consensus building. These and others, including policy proposals, were coded into the “action” category in this study. This was done because interviewees identified them as important bases for their support decisions.

Table IV shows that presidential “action” is indeed used by a great many people as a basis for judgment*.* In fact, in President Obama’s first year in office total action mentions exceeded total results mentions by 39 to 38 percent (not reported in tables). By the fall of 2011, however, this small difference had been reversed, with outcome mentions exceeding action mentions by 41 to 40 percent (not reported in tables). Though the increase is not statistically significant, it does reflect public sensitivity to changes underway that explain the shift to results. For in that short time span, one war had ended and another had moved closer to resolution. Too, impatience with the slow recovery from the Great Recession had increased. More people were willing to hold Obama accountable for the state of the economy. Before this shift, however, citizens had not withheld their judgments just because results were not yet visible to them. They werequite willing to base approval decisions on work in progress. *Theory B is therefore incomplete*. To account for the greatest concentration of citizen judgments (represented in cells2 and 3 of Table IV) the explanation must be extended to include “action” as well as “results.”

Results across Performance Categories

 Although the Action cell attracts fewer mentions than the results cell in the Effectiveness Value row of Table IV, the Acts *column* of Table IV (cells 2, 5, and 8) attracts 57 more mentions than the Outcomes column (cells 3, 6, and 9). Why? The answer becomes clear if the mention count percentages *within each of the four performance domains* addressed in the interviews are examined. The interesting distinctions are in **Tables V and VI,** whichallow direct comparison of the four highest criterion mention percentages in each performance category separately for the second Bush and first Obama terms. These tables show that *the status of results varies across performance categories.* It is not always the most mentioned basis of judgment for either president because (again) *its mention frequency depends on the visibility to citizens of results in specific performance categories at particular times*. Interestingly, “results”isthe most frequently mentioned criterion in just two of the four performance categories: the same two categories for both Bush and Obama: “Good President,” which is based on responses to questions about past presidents, in office before these interviews, and “Economy Handling,” which features the two presidencies in progress at the time of the interviews.

Questions in the first category ask people to reflect on “what makes for a good or successful president” and to give examples of past presidents who they thought did good and poor jobs, respectively. When they are asked to explain their judgments of such *completed* presidential records “results” is the measure of success mentioned most frequently.

Why did the same outcome-centric view shape explanations of current economic assessments for the first-term Obama as well as the second term Bush? Because unlike other presidential performance categories, the economy features regular updates of empirical indicators that (once a president is thought to have been in office long enough) many come to see as *performance outcome measures for the incumbent president*: GDP growth rate, the number of new jobs created and the unemployment rate. Although academics and journalists often point out that these metrics are beyond any president’s direct control, they are widely publicized and often used as results indicators to hold the president accountable for the state of the economy. Moreover, presidential candidates run on economic policy proposals and promises, which invite accountability, and the opposition party routinely blames the incumbent president for bad economic news revealed by these indicators. Thus, many citizens simply take these elite cues and blame presidents for evidence of sluggish GDP, slow job growth and high unemployment. This is true despite the frequent existence of plausible alternative blame targets that, in the first Obama term, included the refusal of partisan congressional opponents to enact key presidential economic proposals, the lingering effects of the Great Recession and ongoing economic instability in Europe. Unsatisfactory economic conditions are the most-mentioned reason for disapproval of Obama’s first-term economic performance, both in these interviews and in contemporary national polls (e.g., Hook, Lippman and King, Jr., June 27, 2012). Misattribution of economic blame is a common citizen mistake, in part attributable to misleading candidate and opposition claims, but also to the fact that many citizens are not well-enough informed to discount such disinformation.

Why does “action” lead the “Incumbent Evaluation” and “War Handling” columns in Tables V and VI as the most-mentioned explanations for approval or disapproval in those performance domains? It does so because the largest proportion of interviewees deemed presidential performance in these categories to be “work in progress” rather than indicative of “results.” Outside of economics many citizens speak of results only when they perceive the approach or presence of recognizable end-states, like the conclusion of a presidential term or the end of a war. This perceptual difference is easiest to capture when people are asked to assess different performance categories. In fact, much was still in progress over most of the Bush and all the Obama interview waves (see wave dates in Table III). Thus, Table VI shows that at the end of his first term two of three of Obama’s evaluated categories—Incumbent evaluation and War Handling featured more Action mentions than Results mentions. But a sizeable plurality thought they had seen quite enough by way of economic outcomes to make results more often mentioned than action when judging his economic performance.

In sum, Brody’s well-known (1991) claim that “results” dominates presidential assessment is just partially right. Results are dominant *only when they are perceived*. The finding that, absent results, many citizens are *no less willing* to rest approval decisions on “action” criteria in any performance category is an important qualification of Brody’s truism, one that merits acknowledgement as a major factor affecting how Americans judge presidents. Over the course of eight years of interviewing action and results “mention” frequencies are often similar, as they are in Table IV. Since major presidential projects are typically at different stages of perceived and actual maturity, this will often be the case. To be sure, some citizens use “traits” or “actions” as information shortcuts (Kuklinski and Quirk, 2000). But evaluative use of “action” understood as a potential precursor to results was not uncommon among this sample of citizen interviewees. Thus, many citizens are basing their approval of presidents on relatively subtle distinctions. If, as Brody contends, citizens deserve credit for fastening on and rewarding results, they also deserve acknowledgment for recognizing and rewarding work in progress. The matrix lets us see that citizen oversight is more nuanced than Brody found or that many political scientists expect (e.g., Lupia and McCubbins, 1998) and potentially more influential and effective than is generally believed.

The Partisan Conundrum

 Brody’s (1991:175) suggestion that voters were “resisting partisan blandishments” is also challenged by the interviews. The first two priorities of citizen judges across the “Good President” and the incumbent performance categories in Tables V and VI are results and action. But for both Bush and Obama term interviewees, the most striking feature of the third and fourth place rows is the *prominence of partisan agreement* (shortened to “partisan” in the tables) as a professed evaluative priority. As noted earlier, partisan agreement is instead a litmus test in which political loyalty is the sole criterion. For Bush (Table V) partisan agreement attracted the third largest cluster of number one mentions in war handling and economy handling, and the fourth largest such cluster for the president’s overall performance. For Obama (Table VI) the same three partisan-dominated performance categories are featured, but *each* appears in row 3, reflecting their emergence as the third-largest concentrations of top-priority “grounds for judgment” mentions in each of the three incumbent performance categories. Does this uptick from Bush to Obama indicate increasing partisan fervor among the electorate? Yes. The percentage of total partisan mentions is significantly higher for Obama than for Bush in two of the three incumbent performance categories*,* particularly in the War Handling column (26% to 11%; p<.05)*.* Obama also attracted 16% to Bush’s 10% in the Incumbent Evaluation category(p<.05). Only in the Economy Handling category did Bush draw a greater percentage of top-ranked partisan mentions than Obama (21% to 19%).

 Partisan Agreement also increased significantly in importance as a driver of general approval in the Obama years. Aggregate (i.e., cross-category) Partisan Agreement comparisons (not reported in tables) show that while President Bush was disapproved of twice as often as he was approved of for partisan reasons (14% to 7%) President Obama was disapproved of *three* *times as often* as he was approved of (75% to 25 %) for partisan reasons. The fact that disapprovals substantially outnumber approvals in all the partisan cells in Tables V and VI shows that it is *opposition partisans, not supportive partisans,* who are *driving partisanship upward in the mention-count rankings of evaluative priorities.* Additional evidence (not in tables) shows that partisans disproportionately approve the performance of the co-partisan president, while disapproving of the opposition president. Thus, Republicans approved of Bush74% to 8%, while Democrats disapproved by a margin of 55% to 13%. Conversely, Democrats approved of Obama’s performance by 90% to 10%, while Republicans disapproved by 88% to 12%.

 Finally, it is significant that partisan agreement, not present among the four most frequently mentioned criteria used to define a “good president” during the second Bush term *emerged in third place on that list during the Obama years,* garnering 9% of the number one mentions. It is one thing to discover that use of the partisan agreement criterion has increased from Bush to Obama. Potentially more worrisome, however, is its emergence as a generic “good president” descriptor. Could this signal the beginning of a shift away from a “valence” perspective emphasizing public interest problem solving to a “power-struggle” model that marginalizes problem solving in favor of ideological fealty and reflexive partisan support or opposition? Some elite Republican criticism of Obama during his first term appeared to fit this description. As noted in the test of Theory B above, “valence” problems (which reflect the “overwhelming consensus as to the goals of government action” Stokes, 1966: 170-173) are clearly still central to presidential assessment. We also see this in presidential “greatness” rankings (Ragsdale, 2009: 27-30). The evidence continues to show majority demand that presidents aim primarily at achieving common goods (security and prosperity). But if evaluative partisanship among citizens continues to grow in importance in subsequent presidential terms, it has the potential to reorder presidential incentives.

Bush-Obama Comparisons

 Next examined are significant changes from Bush to Obama. These men served at different points in time and at different stages in their presidential careers. President Bush was a second-term president during the first four years of interviews, while his successor was serving his first term in office. Table V shows that Bush’s record was *disapproved by majorities or pluralities* in most cells in the three incumbent performance categories. Table VI, on the other hand, shows that the cumulative reaction to President Obama during his initial term was much more favorable. Partisan judgments aside, majorities or pluralities approved of his overall performance, his war management performance and his handling of the economy. What accounts for these dissimilar judgments of the two presidents?

 First, as suggested by the differences in the “Good President” columns in Tables V and VI Bush and Obama faced *different public expectations*. The questions asked people to identify the evidence they used to judge if a president was doing a good job, and then to give examples of past presidents who illustrated the differences between “good” and “poor” presidents that mattered most to interviewees. They were explicitly asked *not* to include the incumbent (see questionnaire in Appendix). But despite this deliberate cueing of the past it is clear that the responses of each group were affected by beliefs about the nature and urgency of curren*t* national issues and by their perceptions of the incumbent president. This is similar to the finding of Kinder, et. al. (1980) that conceptions of an ideal president were “defined, at least in part, by the conspicuous actions and predicaments of the incumbent president” (1980:332). In the present study, these things affected both the appearance of particular criteria and the degree of consensus (percentage magnitudes) among its “Good President” expectations. For example, “results” topped the list for both groups but the Bush group evinced significantly less consensus on this criterion than the Obama group (33% to 52% respectively; p<.05) because of new and more widespread hopes for economic recovery under Obama. The groups agreed that good presidents take action(second-ranked at nearly equal rates of 20% and 22% of #1 “mentions”, respectively)*.* But “assertiveness” was a “Good President” trait only during the Bush term: a likely consequence of the fact that Bush’s assertive leadership style proved to be reassuring to many after 9/11. The last important difference is the already noted emergence of partisan agreement as a “Good President” trait: present under Obama, absent under Bush. Clearly, the “Good President” responses of these interviewees were more indicative of change than of retrospective stability in the climate of expectations.

President Bush

 In the all-important war and economy handling categories the most mentioned grounds used to judge each president were the same: action and results. Because both wars were “works in progress” through the first five years and one was under way for all eight of these interview years the top evaluative criterion in that category across both administrations was action. But Bush had a significantly higher percentage of first place mentions than his successor: 56% versus 44% (p<.05. See the War Handling action cells in Tables V and VI). Why? Obama’s 2012 defeat of John McCain sparked a reduction in the number of those seeing war as a top priority by signaling, for many, a changing of the guard and with it of perceptions of the war’s significance. Obama’s election clearly encouraged the shift of some first place criterion mentions in the War Handling column from war action to the partisan category (mostly disapproving) and to results (mostly approving) as different groups of interviewees took the measure of the new president in the war categories of greatest concern to them. For Bush, however, the War Handling action cell drew the largest proportion of mentions—56%--of any cell in either the Bush or Obama-era tables. This was clearly Bush’s capstone evaluative test, and it was a test that he did not pass. The War Handling column shows greater disapproval percentagesfor him than does the economic column in two most-mentioned incumbent performance cells in Table V.

 For reasons suggested earlier the top economic yardstick for both presidents was results. It is clear that these citizen-judges expected economic improvements from both presidents. The top-ranked results cell of the Economy Handling column in both presidents’ tables shows plurality disapproval for Bush but an impressive level of plurality approval for Obama. Why? First because the context for each president was substantially different. President Bush’s “economic results” reviews are worse because he was nearing the end of his second term with an uninspiring economic record and an emerging economic crisis (the “Great Recession”) on his hands. Bush’s “economic action” percentages were also significantly worse than Obama’s, whose approval scores benefitted from taking bold if controversial action (including the stimulus package and the auto industry bailout): things he did as a new president confronted with a worsening crisis, and things that contributed to an improving economic outlook.

 To conclude on Bush, it is especially noteworthy that many Bush era respondents thought his display of strength in a time of fear important enough to include assertiveness among the qualities that should be displayed by model presidents (Table V). Assertiveness was also the second most-mentioned evaluative target in Bush’s “Incumbent Evaluation” column. This matters greatly to any effort to assess the public’s reaction to this president. The reason is that assertiveness is *the only* source of plurality approval for Bush anywhere in Table V. Nothing else that he did or achieved during his second term rated an approval score among these interviewees as high as the 49% that approved of his assertiveness (second-ranked cell under Incumbent Evaluation). Relatedly, among the minority (30%) who approved of Bush’s overall performance (Table V footnote 1) his assertiveness was the most frequently mentioned reason (38%; not reported in Tables). All these percentages show the enduring appreciation of a sizeable minority for the reassurance provided by President Bush’s forceful post-9/11 posture, even as fear of terrorism receded and unhappiness with his Iraq war policy and the economy increased. But by the time Barack Obama became president the Great Recession had fully supplanted the Terror War as the overriding concern. The belief that “Good Presidents” display assertiveness would not make the top four in the first Obama term.

President Obama

 The largest single mention percentage in Table VI, the Obama table, is results in the “Good President” column(52%) followed in second place by top-ranked results(at 43% of mentions) in the “Economy Handling” column. Action and results in the Obama “Economy Handling” column attracted significantly greater mention percentages than those in the “War Handling” column.

 Given war fatigue and the Great Recession, it is not surprising that the economy, which so directly affects peoples’ daily lives, would trump “war handling” as the focus of their concern. But it is at least ironic that President Bush is the one seen as the “strong leader,” sometimes a reference to a president’s demonstrated willingness to use force. For it is not Bush but President Obama (whose “assertive” attributions were generally anemic, barely achieving a fourth place ranking under “Economy Handling” at just .02 percent of total “mentions” in Table VI) who turned out to be what one expert observer called “one of the most militarily aggressive presidents in decades” (Bergen, April 29, 2012: SR 1). It seems that among Table V’s Bush era interviewees “assertiveness” (second ranked under “Incumbent Evaluation”) was more about the president’s rhetorical stance than his deeds, since most of Bush’s war handling reviews—both action and results--were strongly disapproving. But if it is military results that count Bergen’s itemization of Obama supportive evidence is lengthy and persuasive: decimating Al Qaeda’s leadership, overthrowing Libya’s dictator, increased drone attacks in Pakistan, successful covert wars in Yemen and Somalia, tripling U.S. troops in Afghanistan, and using Navy Seals to remove Osama bin Laden, among other things. Table VI shows majority “War Handling” column approval for Obama in all but the Partisan cell. Yet this study’s interviewees did not seem aware of or particularly concerned with the *scope* of the president’s military initiatives and achievements. Bergen also describes this tendency as a general rule within the mass public but expresses surprise that among both political elites and citizens, neither the left nor the right “regards Obama as the warrior president that he is.”

 How can this apparent inattention to such an extraordinary military record be explained? My hypotheses: The media covered it but their coverage attracted little public notice; the Republican opposition had incentive to downplays it because it threatened their “ownership” of the national security issue; the president’s party had little reason to defend it given public approval of the president’s war handling (and congressional Democrats’ own need to focus mainly on the economy for re-election) and, perhaps most important, *it was not President Obama’s style to trumpet such achievements.* Whereas President Bush used attitude, rhetoric and gesture to call attention to his assertiveness, and did so in a receptive climate of anxiety, his successor (“no drama Obama” to some) faced no comparably intense national security anxiety and was by nature self-contained, matter of fact, and succinct in his portrayals of military decisions, actions, and outcomes.

**Conclusion: The Uses of the Matrix**

Presidential oversight is the core civic duty of American representative democracy, yet citizen performance of this important duty is rarely studied. If the incentives citizens create for presidents are to be clearly understood and as effective as possible, it is necessary that citizens’ oversight performance be routinely assessed; just as presidential performance is routinely assessed. Competent oversight is essential to presidential effectiveness.

Because the matrix organizes citizen testimony using historically important values and research-vetted targets, it is multifaceted enough to surface information that Richard Brody’s correlational method could not; a case in point being the near-equivalent importance of action to results. Applying the matrix to different presidential performance categories also made it clear thataction was the most mentioned criterion only in performance categories where presidential work was still in progress (Tables V and VI). It is to their credit that citizens perceive action that may yield results as a suitable basis for judgment before results become apparent.

Worth noting too is the fact that both interviews and the matrix can be adapted to accommodate special projects, such as the extent to which citizens stay abreast of fast breaking performance news (e.g., President Obama’s handling of Russia’s annexation of Crimea). This is accomplished by recasting standard interview questions to fit the emergency. Responses should reveal if special cases alter the scope and/or intensity of respondents’ focus and learning. Interviews conducted in special circumstances offer potentially revealing comparisons to our non-crisis interviews on matters like oversight motivation and performance.

Finally, the red flag raised earlier by these interviews is important enough to be raised again in this final paragraph. I refer to the emergence of partisan agreement as a high priority requirement for a growing number of citizens; and also to the fact that its mention frequency increased substantially from Bush to Obama. This matters because its continued growth would pose an increased threat to competent citizen oversight and to government’s effectiveness more generally. Increased citizen preoccupation with partisanship reduces the pressure on presidents to emphasize valence problem solving by encouraging presidents to pander to the base. The “power-struggle” model that marginalizes problem solving in favor of ideological fealty and reflexive partisan support or opposition signals that winning is becoming more important than governing and problem solving. That is a recipe for paralysis and decline.

**Endnotes**

1. The standard options for measuring attitudes are closed (forced choice response options are presented to respondents) or open (respondents are asked explicitly to describe their attitudes). Forced-choice measurement dominates. It does so not because open-ended self-reports are invalid, but because of the complexity entailed in coding open-ended questions (see the text and the Lewis-Beck, et.al. 2008: 258 discussion of their “rather tedious” open-ended coding exercise). For present purposes open-ended strategies are not just viable options (Krosnick, et. al. 2005) but uniquely appropriate strategies (Holbrook, et. al. 2001). Not only can open-ended questioning tap the latent construct in focus here, past studies show that they have higher reliability and validity than closed-ended questions (Hurd, 1932; Remmers, et. al. 1923). For a largely uncharted terrain like evaluative attitudes toward presidential performance open-ended questions were worth “the trouble they take to ask and the complexities inherent in the analysis of their answers” (Krosnick, et. al., 2005: 35).
2. Those who dismiss the reliability and validity of such testimony (e.g., Lodge, et.al., 1995: 30-311) contend that it cannot be trusted to indicate how people actually judge presidents, since people may merely “dredge up commonsensical rationalizations” to justify what they have concluded. There are such people but they are in the minority. Eight years of lengthy interviews with 855 citizens made it possible to spot and contextualize them as it became apparent that there were four distinguishable evaluative “types,” three of which articulated authentic evaluative attitudes (Buchanan, 2011). Most of those resembling the “commonsensical rationalizers” described by Lodge, et. al. fit into a group called “Indifferents,” who made up just three percent of this 855 person sample.

**References**

Bailey, Thomas A. 1966. *Presidential Greatness*. New York: Appleton-Century-Crofts.

Bergen, Peter L. April 29, 2012. Warrior In Chief. *New York Times*, p. SR 1

Borgida, Eugene, Christopher M. Federico and John L. Sullivan (eds). 2009. *The Political Psychology of Democratic Citizenship.* New York: Oxford.

Brody, Richard A. 1991. *Assessing the President.* Stanford, CA: Stanford University Press.

Buchanan, Bruce. 2013*. Presidential Power and Accountability: Toward a Presidential Accountability System*. New York: Routledge.

Buchanan, Bruce. 2011. Judging Presidents: Evaluative Types as Presidential Citizens. Paper presented at the 2011 annual meeting of the American Political Science Association, Seattle, Washington, September 1-4, 2001.

Buchanan, W. (1974). Nominal and ordinal bivariate statistics: The practitioner's view. *American Journal of Political Science,* 18 (3), 625-646.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Vote*r. Chicago: University of Chicago Press.

Cohen, Jeffrey E. 1999. Politics and Economic Policy in the United States. 2d Ed. Boston: Houghton-Mifflin.

Caro, Robert A. 2012*. The Years of Lyndon Johnson: The Passage of Power*. New York: Knopf.

Dallek, Robert. 1998. *Flawed Giant: Lyndon Johnson and his Times* 1961-1973.

Delli Carpini, Michael X. and Scott Keeter. 1996. *What Americans Know About Politics and*

*Why it Matters.* New Haven, CT: Yale University Press.

Fenno, Richard E. Jr. 1978. *Home Style: House Members in their Districts.* New York: Scott-Foresman.

Giglio, James. N. 1991. *The Presidency of John F. Kenne*dy. Lawrence, KS: University Press of Kansas.

Gilens, Martin. 1988. Gender and Support for Reagan: A Comprehensive Model of Presidential

Approval. *American Journal of Political Science* 32: 19-49.

Goldstein, Gordon M. 2008. *Lessons in Disaster: McGeorge Bundy and the Path to War in Vietnam*. New York: Times Books, Henry Holt.

Greene, Steven. 2001. The Role of Character Assessments in Presidential Approval. *American Politics Research* 29: 196-210.

Greenberg, Amy S. 2012. *A Wicked War: Polk, Clay, Lincoln, and the 1846 U.S. Invasion of Mexico.* New York: Knopf.

Greenstein, Fred. 2009*. The Presidential Difference*. 3rd.Ed. Princeton, NJ: Princeton University Press.

Gronke, Paul and Brian Newman. 2003. FDR to Clinton, Mueller to ??: A field essay on presidential approval. Political Research Quarterly 56: 501-12.

Harwood, John. June 14, 2010. Criticism Flowing Like Oil, But Obama’s Rating is Steady.

*New York Times*  p. A13. HAJP References

Harwood, John. March 17, 2006. Washington Wire: Losing Ground. *Wall Street Journal*, p. A4

Harwood, John. September 15, 2005. Katrina Erodes Support in U.S. for Iraq War. *Wall Street Journal,*

Hermann, Margaret G. Ed. *Political Psychology.* San Francisco, CA: Jossey-Bass, pp 193-218. Richard R. Lau and Douglas O. Sears. Hillsdale, NJ: L. Erlbaum Associates.

Herring, George C. 2008. *From Colony to Superpower: U.S. Foreign Relations Since 1776.* New York: Oxford University Press.

Hook, Janet, Daniel Lippman and Neil King Jr. June 27, 2012. Obama Clings to a Narrow Lead. *Wall Street Journal*, p. A4.

Holbrook, A.L., J. A. Krosnick, P.S. Visser, W.L. Gardner, and J.T. Cacioppo (2001). Attitudes toward Presidential Candidates and Political Parties: Initial Optimism, Inertial First Impressions, and a Focus on Flaws. *American Journal of Political Science* 45: 930-950.

Howell, Susan E. and Huey L. Perry. 2004. Black Mayors/White Mayors: Explaining their Approval. *Public Opinion Quarterly*, 68:32-57.

Hurd, A.W. 1932. Comparison of Short Answer and Multiple Choice Tests covering Identical Subject Content. *Journal of Educational Psychology* 26: 28-30.

Kinder, Donald R. 1986. Presidential Character Revisited. In *Political Cognition*, edited by Kinder, Donald R. and Susan T. Fiske. 1986. Presidents in the Public Mind. In Margaret G.

Kinder, Donald R., Mark D. Peters, Robert P. Abelson, and Susan T. Fiske. 1980. Presidential Prototypes. *Political Behavior*, 2:315-337.

Knapp, Thomas R. 2009. Percentages: The Most Useful Statistic Ever Invented. http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCEQFjAA&url=http%3A%2F%2Fwww.tomswebpage.net%2Fimages%2FPercentages.doc&ei=qieDUJXQB4Xo2gXVh4HICQ&usg=AFQjCNH6jMVG6y7mg2ax-T26Qa-vQ09OAQ

Kuklinski, James H. and Paul J. Quirk. 2000. Reconsidering the Rational Public: Cognition, Heuristics, and Mass Opinion. In Arthur Lupia, Mathew D. McCubbins and Samuel Popkin, eds. *Elements of Reason: Cognition, Choice and the Bounds of Rationality*. New York: Cambridgle, pp.153-182.

Krosnick, Jon A., Charles M. Judd and Bernd Wittenbrink. 2005. The Measurement of Attitudes. In D. Albrracin, B.T. Johnson and M.P. Zanna (eds.) *The Handbook of Attitudes* (Chapter 2). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Krippendorff, Klaus. 1980. *Content Analysis: An Introduction to its Methodology.* Beverly Hills,CA: Sage Publications.

Kuklinski, James H. and Paul J. Quirk. 2000. Reconsidering the Rational Public: Cognition, Heuristics and Mass opinion. In Arthur Lupia, Mathew D. McCubbins and Samuel L. Popkin, eds. *Elements of Reason: Cognition, Choice, and the Bounds of Rationality.* New York: Cambridge University Press.

Lane, Robert E. 1962*. Political Ideology: Why the American Common Man Believes What He Does.* Glencoe, IL: Free Press.

Lehman, Edward W. 1992. *The Viable Polity*. Philadelphia: Temple University Press.

Lewis-Beck, Michael S., William G. Jacoby, Helmut Norpoth and Herbert F. Weisberg. 2008. *The American Voter Revisited*. Ann Arbor, MI: University of Michigan Press.

Lodge, Milton, Marco R. Steenbergen and Shawn Brau. 1995. The Responsive Voter: Campaign Information and the Dynamics of Candidate Evaluation*. American Political Science Review*, 89: 309-326.

Lupia, Arthur and Matthew McCubbins. 1998. *The Democratic Dilemma*. Cambridge: Cambridge University Press.

Miroff, Bruce. 1976. *Pragmatic Illusions: The Presidential Politics of John F. Kennedy*. New York: McKay.

Mueller, John E. 1973. *War, Presidents, and Public Opinion*. New York: Wiley.

Newman, Brian. 2004. Presidential Traits and Job Approval: Some Aggregate Level Evidence*. Presidential Studies Quarterly* 34: 437-448.

Newman, Brian. 2003. Integrity and Presidential Approval, 1980-2000. *Public Opinion Quarterly* 67: 335-367.

One Year After. March 19, 2004. *New York Times*, p. A18.

Ragsdale, Lyn. 2009. *Vital Statistics on the Presidency,* 3rd ed. Washington, D.C.: CQ Press.

Reeves, Richard. 2001. *President Nixon: Alone in the White House*. New York: Simon and Schuster.

Remmers, H.H., L.E. Marschat, A. Brown and I. Chapman. 1923. An Experimental Study of the Relative Difficulty of True-False, Multiple-Choice, and Incomplete Sentence Types of Examination Questions. *Journal of Educational Psychology* 14: 367-372.

Rich, Frank. 2006. *The Greatest Story Ever Sold: The Decline and Fall of Truth from 9/11 to*

*Katrina.* New York: Penguin.

Rudalevige, Andrew. 2005. *The New Imperial Presidency*. Ann Arbor, MI: University of Michigan Press.

Savage, Charlie. 2007. *Takeover: The Return of the Imperial Presidency and the Subversion of American Democracy.* New York: Little, Brown.

Simonton, D. K. (2008). Presidential greatness and its socio-psychological significance: Individual or situation? Performance or attribution? In C. Hoyt, G. R. Goethals, & D. Forsyth (Eds.), *Leadership at the crossroads: Vol. 1. Psychology and leadership* (pp. 132-148). Westport, CT: Praeger.

Skowronek, Stephen. 1993. *The Politics Presidents Make*. Cambridge: Harvard University Press.

Stokes, Donald E. 1966. Spatial Models of Party Competition. In Angus Campbell, Phillip Converse, Warren Miller and Donald Stokes, *Elections and the Political Order*. New York: Wiley, pp. 161-179.

 Stokes, Donald E. 1992. 'Valence Politics', in Dennis Kavanagh, ed., Electoral Politics (Oxford: Clarendon Press, 1992).

Trochim, William M. 2008. Measurement: Construct Validity and External Validity. In *The Research Methods Knowledge Base* 2nd Edition at URL<http://www.socialresearchmethods.net/kb/>

Table I

**The “Values and Targets” Hypothesis**

 Traits Acts Outcomes

|  |  |  |  |
| --- | --- | --- | --- |
| Effectiveness | Assertiveness 1 | Action2 | Results3 |
| Morality | Virtue4 | Moral Action5 | Ethical Results6 |
| Prudence | Caution7 | Prudent Action8 | Results prudently achieved9  |

Figure I

 **Coding Examples for Each of Nine Table 1 Matrix Cells**

(I approve/disapprove because…)

**Effectiveness Row Cells**

 1 Assertiveness “President Bush is a decisive person.”

 “I don’t think President Obama is tough enough with Republicans”

 2 Action “President Obama supported extension of the Bush tax cuts.”

“President Bush should not be advocating limits on stem cell research.”

 3 Results “President Bush helped eradicate AIDS in Africa.”

“President Obama said his stimulus plan would get unemployment below 10 percent by now, but it hasn’t.”

**Morality Row Cells**

4 Virtue “It was generous of Obama to acknowledge Speaker Boehner’s rise from modest beginnings in his 2014 State of the Union Address.”

 “It struck me as hypocritical for Bush to brandish his godliness while savagely attacking Kerry in his campaign ads.”

5Moral Action “No Child Left Behind was a worthy effort by President Bush to close the gap between black and white, rich and poor.”

“I do not believe it is right for President Obama to use drones because they threaten civilians as well as terrorists.”

6 Ethical Results “I think Bush did the right thing when he ignored conservative ideology to support the TARP bailout. It prevented another Great Depression.”

 “It was wrong for Obama to tell us we could keep our health insurance coverage just to win support for the ACA, then to cancel our coverage after the law was on the books.”

**Prudence Row Cells**

7 Caution ”Obama is a careful, thoughtful guy.”

 “Bush is an instinct player.”

8 Prudent Action “President Obama took great pains to make a well-grounded decision before sending more troops to Afghanistan.”

“The way Bush rushed us into Iraq seemed reckless to me.”

9 Results Prudently Achieved

 “Obama was smart to accept Putin’s offer to help get Syria to give up its chemical weapons. Better that way than with air strikes!”

 “It bothers me that Bush so often rolls the dice on a risky move instead of proceeding more carefully. Iraq was the best example, but don’t forget that giant tax cut! What did either move get us?”

Table II

**Aggregate Sample Demographics 2005-2012**

855 Interviewees compared to recent national samples (in parentheses)

 Percentage Number Missing

 (rounded) Data

Gender1

 Male (48%) 50 431

 Female (52%) 50 424

Party Identification2 8

 Dem (32%) 35 300

 Rep (27%) 30 257

 Ind (35%) 34 286

 Other 1 4

Education1 12

 HS (31%) 29 247

 College (64%) 71 596

Age3  2

 18-39 (31%) 34 290

 40-59 (41%) 37 318

 60+ (29%) 29 245

Typology4  2

 Delegator (56%) 59 502

 Ideologue (8%) 11 98

 Instrumental (33%) 27 228

 Indifferent (3%) 3 25

Vote in 2004 (N=549)5

 Yes 71 393

 No 29 157

Vote in 2008 (N=473)5

 Yes 85 468

 No 15 81

Vote Opposing Party (N=301)5, 6

 Never 54 189

 Sometimes 43 149

 Often 3 10

1Gender and education percentages in parentheses (rounded) are from a national sample of 1509 adults conducted by Princeton Survey Research International for the Pew Research Center for the People and the Press, 02/05/11, 2007.

2Party ID percentages are from the Pew Research Center, January 5-9, 2011.

3Age percentages are estimates from the February 2007 Pew poll cited above. Estimates are necessary because Pew survey data are not precisely comparable to interview sample data due to sampling of slightly different age ranges.

4Typology description and national distribution estimates of types from Buchanan, 2007.

5Questions added in mid-project.

6Excludes Independents and Other.

Table III

**Interviews by Waves\*>**

|  |  |
| --- | --- |
| *Excel # Location*  | *Wave Years*  |
| 1 – 45 | Spring 2006 |
| 46 – 129 | Spring 2005# |
| 130 – 180 | Fall 2006 |
| 181 – 300 | Fall 2007 |
| 301 – 385 | Fall 2008 |
| 386 – 463 | Fall 2009 |
| 464 – 504 | Spring 2009 |
| 505 – 598 | Summer 2009 NY Web Interviews\*\* |
| 599 – 661 | Fall 2010 |
| 662 – 702  | Spring 2011 |
| 703 – 779 | Fall 2011  |
| 780 – 799  | Spring 2012 |
| 800 – 855  | Fall 2012 |

\*Dissimilar wave sizes and temporal spacing due to exigencies of class size, recruitment and scheduling

>Total number of interviews=855

#Spring 2005 preceded by spring 2006 due to data entry error

\*\*Summer 2009 New York interviews (statewide random sample) conducted to check for regional distortion. No systematic differences between Southwest and Northeast interview responses were detected

Table IV

**Test of Values and Targets Hypothesis**

Theory A

855 Interviewees

2005-2012

#1 Evaluative criterion “mentions”

for each of nine cells summed across

four presidential performance categories

81% fit to Matrix1

|  |
| --- |
|  Presidential Presidential Presidential  Traits Acts Outcomes  |
| Effectiveness.MoralityPrudence |  Assertiveness ***216***1(7%)2 | Action***1,098***2 (39%) | Results***2,519 (89%)*** ***1,205***3  (42%) |
| Virtue***98***4 (3%) | Moral Action***98***5 (3%) | Ethical Results ***201 (7%)******5***6 (.1%) |
| Caution***5***7 (.1%) | Prudent Action***91***8 (3%) | Safe Results ***116 (4%)******20***9 (.7%) |

***319 (11%)*** ***1,287 (45%)***  ***1,230 (43%)***  ***2,836***

1 2,836 #1mentions of a total of 3,503 #1 “mentions” fit the matrix. 2,836/3,503 = 81% of mentions fit matrix. 667 additional #1 mentions, categorized as follows, did not fit matrix: Partisanship = 527 (15%); Other = 140 (4%). All percentages in the table are rounded.

2 Cell count as a percentage of total #1 matrix fitting mentions.

Table V

**George W. Bush**

**Criteria Rankings by #1 Mention Percentages**

**“Good President” and Three Incumbent Performance Categories**

385 Interviewees

2005 – 2008

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Rankings** | **Good President** | **Incumbent****Evaluation1** | **War****Handling2** | **Economy****Handling3** |
| 1 | Results (33%) | Action (35%) Approve 29% Disapprove 52% Mixed 19% | Action (56%) Approve 37% Disapprove 47% Mixed 16% | Results (45%) Approve 25% Disapprove 44% Mixed 26% |
| 2 | Action (20%) | Assertive (23%) Approve 49% Disapprove 40% Mixed 10% | Results (20%) Approve 15% Disapprove 66% Mixed 19% | Action (22%) Approve 24% Disapprove 56% Mixed 20% |
| 3 | Assertive (18%) | Results (18%) Approve 18% Disapprove 61% Mixed 19% | Partisan (11%) Approve 20% Disapprove 76% Mixed 5% | Partisan (21%) Approve 24% Disapprove 59% Mixed 17% |
| 4 | Virtuous (10%) | Partisan (10%) Approve 21% Disapprove 74% Mixed 5% | Prudent Action (5%) Approve 10% Disapprove 75%  Mixed 15% | Other (6%) Approve NR Disapprove NR Mixed NR |

1 Bush overall approval: Yes 30% No 54% Mixed 16%

2 Bush war handling approval: Yes 28% No 55% Mixed 16%

3 Bush economy handling approval: Yes 25% No 50% Mixed 25%

Table VI

**Barack Obama**

**Criteria Rankings by #1 Mention Percentages**

**“Good President” and Three Incumbent Performance Categories**

469 Interviewees

2009-2012

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Rankings** | **Good President** | **Incumbent****Evaluation1** | **War****Handling2** | **Economy****Handling3** |
| 1 | Results (52%) |  Action (38%) Approve 63% Disapprove 16% Mixed 21% |  Action (29%) Approve 52% Disapprove 27% Mixed 21% |  Results (43%) Approve 46% Disapprove 26% Mixed 28% |
| 2 | Action (22%) |  Results (32%) Approve 35 % Disapprove 31% Mixed 34% |  Results (27%) Approve 53% Disapprove 32% Mixed 16% |  Action (31%) Approve 78% Disapprove 22% Mixed <1% |
| 3 | Partisan (9%) |  Partisan (16%) Approve 12% Disapprove 78% Mixed 10% |  Partisan (26%) Approve 31% Disapprove 40% Mixed 29% |  Partisan (19%) Approve 13% Disapprove 70% Mixed 17% |
| 4 | Moral Action (7%) |  Moral Action (4%) Approve 59% Disapprove 27% Mixed 14% |  Prudent Action (10%) Approve 35% Disapprove 38%  Mixed 27% |  Assertiveness (.02%) Approve 33% Disapprove 22% Mixed 44% |

1 Obama overall approval: Yes 44% No 32% Mixed 24%

2 Obama war handling approval: Yes 42% No 17% Mixed 42%

3 Obama economy handling approval: Yes 35% No 39% Mixed 26%

. **Appendix**

**INTERVIEW PROTOCOL**

**Fall, 2009-Fall, 2012**

**Thank you for speaking with me! First, we’ll talk about US presidents generally. Then we’ll talk specifically about the current president, Barack Obama. Finally, we’ll talk about two areas of national interest: The wars in Afghanistan and Iraq and the economy.**

**Let’s start with some general questions about what makes for a good or a successful president**

1. First, what kind of evidence or cues do you personally look for to help you decide if a president is doing a good job?

(Prompt for detail, ask them to clarify if something they say is not clear.)

 Ex: Anything else? Can you say more? Please go on.

 Could you explain what you mean by \_\_\_\_\_ ?

2. Can you give me an example of a past president who you believe did a particularly good job? Why him in particular? Anything else?

3. Now can you give me an example of a past president who you believe did a particularly poor job? Why him in particular? Anything else?

**Now let’s talk specifically about President Obama**

1. Do you approve or disapprove of the way Barak Obama is handling his job as president?

Why or why not? Can you say more? Anything else?

2. What do you see as President Obama’s main strengths as a leader? Anything else?

3. What do you see as President Obama’s main weaknesses as a leader?

4. Why do you think Barack Obama was elected president?

5. From what you know of them, what do you think of Obama’s goals for his term?

**Next, some questions about Iraq and Afghanistan**

1. First, what is your understanding of president Obama’s plan for dealing with the Iraq War?
2. Next, what is your understanding of president Obama’s plan for dealing with the war in Afghanistan?

3. Do you approve or disapprove of the way Barack Obama is handling the situations in Afghanistan? Why or why not? Can you say more?

4. Do you approve or disapprove of the way Barack Obama is handling the situations in Iraq? Why or why not? Can you say more?

1. What would you advise President Obama to do next in Iraq? What changes would you make in his policy (if any)?

6. What would you advise President Obama to do next in Afghanistan? What changes would you make in his policy (if any)?

**Now a few questions about the economy**

1. Do you approve or disapprove of the way Barack Obama is handling the economy? Why or why not? Can you say more?

2. How would you describe president Obama’s economic policy? Anything else?

3. What do you think is the most important economic problem facing the country right now? Are there other economic problems you see as almost as important?

4. As you see it, what can a president do to make the economy work better? Anything else?

5. Can you give me an example of a past president who you believe did a particularly good job of handling the economy? Why him in particular?

6. What would you advise president Obama to do next to improve the economy? What changes would you make in his policy, if any?

**Next, a few brief questions about you**

These questions will be used only to ensure that we have interviewed a representative sample of individuals.

Would you describe yourself as:

 A Democrat, a Republican, or an Independent? (other?)

 A conservative, a liberal or a moderate? (other?)

Which sentence best describes your educational background?

 Attended or graduated from high school

 Attended or graduated from college

Which of the following age ranges includes you?

 18-39

 40-59

 60 and above

What is, or was, your principal occupation?

**Lastly, a few questions about voting**

Did you vote in the 2004 presidential electi on? (yes or no)

Did you vote in the 2008 presidential election? (yes or no)

(If you are a Republican or Democrat) How often have you voted for a presidential candidate of the opposing party: never, sometimes, or frequently?

**CODING SHEET**

**INTERVIEW CODING SHEET** (Fall, 2012)

Within Parts A, B, C and D below, use answers to all Q's to figure out **what criterion/criteria** **interviewees use to decide** **if a president is doing a good job** in that part. If more than one criterion is used in any part, indicate your estimate of the rank order of importance of the criteria to the interviewee.

**A. "Good President"** Questions B. **Obama "General Evaluation"** Questions

 OE12 **1** aprv **2**  dis **3** mixd

 **Why approve/disapprove?**

GP1 \_\_\_\_\_ assertiveness \_\_\_\_\_ OE13

GP2 \_\_\_\_\_ purposive action \_\_\_\_\_ OE14

GP3 \_\_\_\_\_ expected results \_\_\_\_\_ OE 15

GP4 \_\_\_\_\_ virtue \_\_\_\_\_ OE 16

GPS \_\_\_\_\_ moral action \_\_\_\_\_ OE 17

GP6 \_\_\_\_\_ untainted results \_\_\_\_\_ OE18

GP7 \_\_\_\_\_ caution \_\_\_\_\_ OE19

GPS \_\_\_\_\_ prudent action \_\_\_\_\_ OE20

GP9 \_\_\_\_\_ safe results \_\_\_\_\_ OE21

GP10 \_\_\_\_\_ policy/ideology/partisan ID \_\_\_\_\_ OE22

GP11 \_\_\_\_\_ other (describe, give examples/quotes) \_\_\_\_\_ OE23

C. **War** Questions D. **Economy** Questions

 WQ24 **1** aprv **2**  dis **3** mixd EC36 **1** aprv **2**  dis **3** mixd

 **Why approve/disapprove?** **Why approve/disapprove?**

WQ25 \_\_\_\_\_ assertiveness \_\_\_\_\_ EC37

WQ26 \_\_\_\_\_ purposive action \_\_\_\_\_ EC38

WQ27 \_\_\_\_\_ expected results \_\_\_\_\_ EC39

WQ28 \_\_\_\_\_ virtue \_\_\_\_\_ EC40

WQ29 \_\_\_\_\_ moral action \_\_\_\_\_ EC41

WQ30 \_\_\_\_\_ untainted results \_\_\_\_\_ EC42

WQ31 \_\_\_\_\_ caution \_\_\_\_\_ EC43

WQ32 \_\_\_\_\_ prudent action \_\_\_\_\_ EC44

WQ33 \_\_\_\_\_ safe results \_\_\_\_\_ EC45

WQ34 \_\_\_\_\_ policy/ideology/partisan ID \_\_\_\_\_ EC46

WQ35 \_\_\_\_\_ other (describe, give examples/ quotes) \_\_\_\_\_ EC47

E. **Demographics (circle)**

DE48 **Gender** Male (1) Female (2)

DE49 **Party ID** Rep (1) Dem (2) Ind (3)

DE50 **Education** HS (1) College (2)

DE51 **Age** 18-39 (1) 40-59 (2) 60+ (3)

**Occupation** (write in) ­­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

TY52 **Typology:** **Delegator (1) Ideologue (2) Instrumental (3) Indifferent (4)**

VO53 Vote in 2004? Y(1) N(2) VO54 Vote in 2008? Y(1) N(2)

VO55 Vote for opposition? Never (1) Sometimes (2) Often (3)Note: For A, B, C, and D above, please write brief reason for coding choices as reminder in margin.