Measuring Political Knowledge: Putting First Things First

Michael X. Delli Carpini; Scott Keeter

American Journal of Political Science is currently published by Midwest Political Science Association.

Measuring Political Knowledge: Putting First Things First*

Michael X. Delli Carpini, *Department of Political Science, Barnard College, Columbia University*
Scott Keeter, *Department of Political Science and Survey Research Laboratory, Virginia Commonwealth University*

Research in political behavior has increasingly turned to the cognitions underlying attitudes. The simplest of these cognitions are political facts—the bits of information about politics that citizens hold. While other key concepts in political science—partisanship, trust, tolerance—have widely used (if still controversial) measures that facilitate comparisons across time and among studies, the discipline has no generally accepted measure of the public's level of political information. This paper describes the development and testing of survey-based measures of political knowledge, with special attention to the existing items on the National Election Study surveys. In so doing, it illustrates the use of a variety of techniques for item analysis and scale construction. We also present a recommended five-item knowledge index.

From Concept to Measurement

While "attitudes" remain central to the study of political behavior, increasing attention is being given to the cognitions underlying attitudes. The simplest of these cognitions are political facts—the various bits of information about politics that citizens hold. This paper describes the development of valid and reliable indicators of citizens' general political knowledge about national politics. In so doing, it also explicates larger issues of scale development and measurement.

The study of factual political knowledge has not, until recently, been especially extensive or sophisticated. Early Gallup surveys did regularly test knowledge of politics, and Berelson, Lazarsfeld, and McPhee's (1954) study of the 1948 presidential election included a large number of factual political questions in its citizen survey. However, once the first

public opinion and voting studies dispatched the romantic image of the "informed citizen," few scholars thought the matter worth further exploration. Political knowledge is rarely measured on contemporary public interest polls, and the National Election Study (NES) surveys from 1964 to 1980 asked very few knowledge questions. What more can be said about a dead horse, after all?

Political scientists did not abandon the study of cognitions, of course—far from it. Great attention has been paid to the matter of "belief systems" and "cognitive complexity" in public opinion. The debate over the existence, measurement, evolution, and implications of the public's "levels of conceptualization" has occupied as much space in the professional journals as any single controversy in the discipline.

But a "belief system" is not just the *organization* of cognitions, which was at the heart of the controversy. Luskin's (1987) paper on the measurement of "political sophistication" provides a useful definition: A *political belief system* consists of "a person's political cognitions, together with those with which they are constrained" (858). Political belief systems vary on three dimensions: size (the number of cognitions), range (the coverage of the political universe), and organization (constraint). "Sophistication is the conjunction of these dimensions," writes Luskin (1987, 860). The debate over "levels of conceptualization," ideology, or whatever one wishes to call it focused mainly on constraint or organization without much consideration of the other dimensions. This was, one might say, putting the cart before the dead horse.

Recent research, much of it drawing upon the concepts and methods of cognitive psychology, has "rediscovered" political knowledge, both as a causal or intermediary variable and as a phenomenon to be explained in its own right. A common conclusion in an increasing number of studies is that factual knowledge is the best single indicator of sophistication and its related concepts of "expertise," "awareness," "political engagement," and even "media exposure" (Luskin 1987; Lodge, McGraw, and Stroh 1989; McGraw and Pinney 1990; Smith 1989; Krosnick and Milburn 1990; Fiske, Lau, and Smith 1990; Zaller 1990, 1992; Price and Zaller 1990).

This renewed analytic and empirical interest raises important, yet largely ignored, questions of measurement. Other key concepts in political science (e.g., partisanship, trust, tolerance) have widely used (if still controversial) measures that facilitate comparisons across time and among studies. But the discipline has no generally accepted measure of the public's knowledge about national politics. The NES surveys have a few direct, and several indirect, measures of political information. While previous research has shown that these measures perform comparatively

well for a variety of purposes (Zaller 1986), their use remains a haphazard approach to quantifying political knowledge.

In this paper, we describe the development and evaluation of several measures of national political knowledge. The first two sections describe our efforts to identify good survey items and to examine the dimensionality of political knowledge. To do so, we constructed a comprehensive political knowledge test and administered it by telephone to a national probability sample of adults. The next section evaluates various knowledge items found on the NES surveys. The final section addresses the validity of survey-based measures of political knowledge.

Issues in Test Construction

Test construction is a multistep process that begins with the development of individual items. The items should be pretested and then administered to a sample that is representative of the population on which the resulting instrument is to be used.[1] Accordingly, we designed a battery of political knowledge questions, which in combination with attitudinal, behavioral, and demographic measures, would constitute a comprehensive survey instrument.[2]

Content Validity: What Should People Know about Politics?

The selection of specific test items is a crucial step, the success of which determines the *content validity* of the eventual measure. Content validity is variously defined as "the extent to which a set of items taps the content of some domain of interest" (Zeller and Carmines 1980, 78), "the adequacy with which important content has been sampled and . . . the adequacy with which the content has been cast in the form of test items" (Nunnally 1978, 93), and "how well the tasks in the test do in fact match the definition of [the] domain" (Thorndike 1982, 184). As Nunnally stresses, the selection of content ultimately depends upon values; it is not amenable to a mechanical or mathematical solution. The notion of "sampling" is applied metaphorically, at best. One cannot easily define the "sampling frame" of political facts. How, then, does one define a relevant domain from which to select the specific questions for a "national civics test"?

---

[1]For a readable general introduction to the construction of summated rating scales, see Spector (1992).

[2]The measures of attitudes and behaviors were included for two reasons: (1) to help establish the construct validity of knowledge measures and (2) to help explain individual variations in levels of knowledge.

Those of us who teach political science—especially introductory courses in U.S. politics—have dealt with this question each time we assembled a syllabus. And so we drew on our own experience as instructors and on our exposure to the canon as expressed in recent textbooks. We also examined high school texts and works by democratic theorists and others who wrote on citizenship and civic education. And we reviewed several decades of public opinion surveys to find questions that tap factual knowledge.

We found considerable agreement that the core should be, in Barber's words, "what government is and does" (1969, 38). Neuman operationalizes the notion of "what government *is*" as "the basic structure of government—its basic values, such as citizen participation, majority rule, separation of powers, civil liberties, and its basic elements, such as the two-party system, the two houses of Congress, the role of the judiciary, and the organization of the cabinet" (1986, 186). A citizen's knowledge of what the government *does* is well described by Berelson, Lazersfeld, and McPhee: "The democratic citizen is expected to be well-informed about political affairs. He is supposed to know what the issues are, what their history is, what the relevant facts are, what alternatives are proposed, what the party stands for, what the likely consequences are" (1954, 308).

In addition to these two broad areas, several others emerged as significant. In a republican form of government, selecting public officials is often a citizen's most significant political power. Therefore, knowledge about political leaders, political parties, and contemporary political alignments seems essential to effective citizenship. Finally, knowledge of "related fields" such as political history and political economy provides an important context for comprehending many elements of current politics.

Experts are typically used both to define relevant domains and to assess the appropriateness of specific items. We sought such "expert judgment" (Crocker and Algina 1986, 68) through a mail survey of a random sample of 111 U.S. political scientists in early 1989.[3] The questionnaire asked their views on both the general topics and specific facts that an average citizen should know. Of eight topic areas presented in the survey, only one was regarded as "essential" by a majority of the respondents: "institutions and processes" of government (65%). When "essential" and "important" responses were combined, the topic areas of institutions and processes (89%), issues and policies (82%), history

---

[3]Questionnaires were mailed to a random sample of 233 members of the American Political Science Association. A total of 111 usable replies were received; 31 packets were undeliverable due to bad addresses, making the adjusted response rate 55% (111/202).

(77%), and current political alignments (67%) all were mentioned by two-thirds or more. We also presented a lengthy list of specific facts and asked the respondents to check those they thought were important for citizens to know (respondents could also suggest their own items). Our survey of political knowledge conformed to the political scientists' views of the most important topics and included eight of the top 10 specific items they endorsed.[4]

Item Format and the Problem of Guessing

Specialists in test construction disagree about the best item format for ability testing, though many suggest that a variety of formats is necessary to cover an appropriate range of content. Multiple choice items have many proponents (Nunnally 1978, 259–60). True/false and either/or formats provide easy opportunities to guess, a phenomenon that lowers the reliability (and hence, the validity) of items.

The mode of administration affects the choice of item format. For example, multiple choice items with several response options can be awkward and time-consuming in telephone surveys. Consequently, we used fewer such items than we preferred and limited the number of choices to three. Several items used a "free response" format; some of these used precoded alternatives to assist the interviewers in recording responses. However, we feared these questions would be especially damaging to respondent morale. This, along with the potential error introduced by interviewer judgment, led us to limit the use of this format despite its advantages.

Some specialists recommend telling achievement test subjects to guess (Nunnally 1978, 648–49) in order to minimize biases arising from differential propensities to guess. We concluded, however, that the unreliability introduced by guessing (especially on items with only two response options) was the more serious problem. Accordingly, we discouraged guessing by explaining to respondents that "many people don't know the answers to these questions so if there are some you don't know just tell me and we'll go on."

Pilot Survey Length

The number of items tested should be much greater than would be included on a final version of a scale (Allen and Yen 1979, 118–19). Nunnally (1978, 261–62) suggests that there be one-and-a-half to two times as many items as the final instrument is projected to have and that there be five to 10 times as many respondents as there are test items.

[4]The questionnaire and other survey details are available from the authors.

Since we expected to develop several short scales ranging from five to 20 items, our survey instrument contained 42 factual test items; the sample size was 610.

Pilot Testing to Determine the Structure of Knowledge and the Individual Performance of Items

The Survey of Political Knowledge was conducted by telephone 21 March–8 May 1989 with a national sample. Methodological details are provided in the appendix. As discussed earlier, this survey served two purposes: to determine the dimensionality of political knowledge and to judge the relative effectiveness of different items for measuring political knowledge.

The Structure of Knowledge: Unidimensional or Multidimensional?

A key measurement issue is whether the concept being measured is unidimensional or multidimensional. A unidimensional concept can be measured by a single index or scale composed of homogeneous indicators. If the concept is multidimensional, a valid measure must include indicators of the various dimensions, and the indicators must be combined or weighted in a meaningful way so as to reflect the appropriate relevance of the dimensions to the overall concept (Nunnally 1978, 274).

Whether political knowledge is unidimensional or multidimensional depends partly on whether citizens are "generalists" or "specialists." It also depends on whether the opportunity to learn about politics varies by topic and across groups of citizens. Prior research on the dimensionality of political knowledge has yielded mixed results, with some scholars arguing that the structure is fundamentally unidimensional (Neuman 1986; Lau and Erber 1986; Zaller 1986; Smith 1989) and others that it is multidimensional (Iyengar 1986; Owen and Stewart 1987; Bennett 1990; Krosnick 1990).

As described above, we designed our survey to take account of several different types of political knowledge. Using both exploratory and confirmatory factor analyses, we found that one could tease out theoretically meaningful dimensions, but that a one-dimensional model adequately represented the structure in the sample data. For example, using LISREL we tested the relative fit of several different hypothesized models.[5] For a one-factor model, the coefficient of determination was .93,

---

[5]LISREL is a statistical program for analyzing the linear structural relationships among a set of variables through confirmatory factor analysis. With LISREL, the *actual* interrelationships the variables (based upon a covariance or correlation matrix) is compared to a *hypothesized* set of interrelationships. If the difference between the hypothesized model

and the ratio of chi-square to degrees of freedom was 2.1, both indicating a strong fit with the data. On the other hand, the *best* fit we could achieve—a five-factor model—had a coefficient of determination of .98 and a chi-square ratio of 1.5.[6] By the standards usually applied in LISREL analyses, these improvements are marginal. In addition, high interfactor correlations among the five factors (ranging from .52 to .94 with a median of .785) suggest that individuals knowledgeable about one aspect of politics were apt to be knowledgeable about others.[7]

Tests of construct validity, in which we regressed various knowledge indexes (based on hypothesized subdomains) onto a set of demographic and psychological predictor variables, confirmed the conclusions drawn from LISREL.[8] Gender, age, and strength of partisanship did show some substantively significant variation across the different subdomains. For example, women were generally less politically informed than men, but this was less true for issues like abortion or women's suffrage. Young adults were less informed about most aspects of politics than were older adults, but not about political institutions and processes, which is regularly taught in the schools. And strong partisans were more likely than others to know the partisanship of Nixon, FDR, and Truman, and which party controls Congress. Nonetheless, the dominant pattern was consistency in the size and direction of the relationships across scales and indicators.

We concluded from this analysis that measures of national political knowledge in one domain can provide reasonably good—though not ideal—measures of overall knowledge about national politics.[9] This is

---

and actual data is small, then the model is considered a plausible simplification of the latent structure underlying the data. If, on the other hand, the differences are great, then the model is deemed less plausible or implausible. Because of our small sample size, our analyses used a representative subset of 18 variables (as recommended when using LISREL). Other analyses using composite variables built from all questions produced similar results (Delli Carpini and Keeter 1993).

[6] The five-factor model distinguished among knowledge of "substantive issues," "institutions and processes," "gender-specific issues," "public figures," and "political parties."

[7] The lowest interfactor correlations were between "gender issues" and "party politics" (.52) and between "gender issues" and "institutions and processes" (.70).

[8] Further discussion of construct validity appears below. For an illuminating example in the present context, see Zeller and Carmines (1980, 91–97).

[9] This conclusion is limited, however, to the arena of national politics. Analyses of data from several state and local surveys we conducted suggest that knowledge of these two levels of government is structurally distinct from national knowledge (Delli Carpini and Keeter 1993). Accordingly, an optimal measure of national political knowledge may not be optimal for discriminating among individuals on their sophistication about state or local politics.

good news for scholars who use the NES surveys, since they include relatively few direct knowledge measures. As we demonstrate below, acceptable measures of knowledge can be cobbled together from the available items (in part because the items themselves usually cover several important domains) and can provide comparable discrimination over time (see, e.g., Zaller 1991).

We caution, however, that there are undoubtedly limits to this happy state of affairs. For example, the notion of "issue specialists" may still apply to narrowly defined domains (e.g., detailed knowledge about social security benefits or tax codes). Knowledge of these domains may, therefore, be quite weakly correlated with general political knowledge. Furthermore, even when domain-specific knowledge is highly correlated with general political knowledge, researchers interested in the former may still benefit from the more specific knowledge measure.[10]

Finding the Best Measures: Item Analysis

Most tests of ability are composed of several—sometimes a great many—individual items. The items are often thought to be "sampled from" the large domain of items related to the trait under study.[11] As noted earlier, this sampling process is not random, since the population of items cannot be specified. And clearly some items are better than others. Under the assumptions of *classical test theory* (also called *true score theory*), an observed score on a test (X) is equal to that true score (T) plus measurement error (e). All things being equal, a better measure is one with less measurement error. A measure with a relatively high proportion of true variance to observed variance (which includes error variance) is said to be a more *reliable* measure (Zeller and Carmines 1980, 13).

One common and useful method for selecting individual items is to choose those with the highest correlations with the total test score.[12] If the items in the test can be assumed to measure the trait under study

---

[10]For a related argument regarding contemporary survey research practice, see Yankelovich (1991).

[11]Throughout this paper, we use terminology common to the literature on psychological testing—often referring to knowledge as an "ability" or a "trait." However, we use these terms in their narrowest sense, preferring to think of knowledge as a resource that can be built up over time. Intellectual ability may help people understand and remember political information, but the type of factual knowledge necessary for good citizenship (and that we probe in our surveys) can be comprehended and retained by individuals of modest cognitive ability.

[12]The item-total correlations are "corrected" by removing the item from the total score before computing the correlation.

(and only that trait), then the relative item-total correlations reflect how much measurement error is present in each item. The most reliable over-all set of items (shorter than the original test) will be those items with the highest item-total correlations. One caveat in the use of item-total correlation is the fact that items with highly skewed distributions (i.e., very difficult or very easy items) will have attenuated correlations with the total test score as a result of how the product-moment correlation coefficient is calculated (Nunnally 1978, 140–46). Thus, choosing items solely on item-total correlation would lead to the omission of very hard and very easy items, even though they might be highly content-valid and reliable.

An "automated" method of item selection based on item-total correlations is stepwise multiple regression, in which the total test score is regressed on the component items. The first item selected will be the one with the highest item-total correlation, but subsequent items chosen will not necessarily be those with the highest zero-order correlations. In essence, the procedure seeks out those items that best resolve the remaining variance. In so doing, inherent multidimensionality in the set of items is at least partially accounted for. Items highly correlated with those already in the model will usually not be chosen, since they provide little additional predictive value. This method is completely atheoretical, however, and should be used with caution (see Lewis-Beck 1978, esp. 218–24). It is also characterized by diminishing returns following the selection of about one-fifth to one-fourth of the items. Discrimination among the remaining items is poor.

Another common criterion for selecting items is *item difficulty,* which is usually defined as the proportion correct for the item (the p value). A test composed of items with a p value of .5 will have a larger variance than if items of greater or lesser difficulty are included. This may be an attractive quality for some testing purposes, especially those where discrimination among subjects in the middle range of ability is desired. However, better discrimination among subjects across a range of ability levels is achieved by using items of varying levels of difficulty. No standard for choosing items based on difficulty exists, and some specialists argue that item difficulty should not be a consideration, except insofar as extreme p values (such as .1 or .9) should be avoided (Nunnally 1978, 270–74). A rule of thumb endorsed by some experts suggests choosing items with p values ranging from about .3 to .7 and averaging about .5 (Allen and Yen 1979, 120–24). Since guessing can artificially inflate p values, one might want to increase the target range of p values to compensate.

Item-total correlation and item difficulty are the two most common

statistics used in item analysis, and both are closely related to classical test theory. An alternative model finding increasing acceptance in the construction and analysis of ability tests is *item response theory* (IRT), which is based upon the broader notion of *latent trait theory*.[13] Item response theory holds that a subject's test performance can be explained by his or her level of a hypothesized latent trait or ability and that the probability of correctly answering a given test item can be expressed as a monotonically increasing function called an *item characteristic curve* (ICC). If appropriate assumptions are met, techniques utilizing item response theory can yield estimates of both the discriminating power and the difficulty of a test item—independent of the particular sample on which they are tested.[14] Conversely, using a test composed of items for which the discrimination and difficulty parameters have been established, the latent ability of the examinees can be estimated—again, regardless of the average level of ability of the sample being tested. The quality of *sample invariance* is one of the most appealing aspects of IRT and has led to its use in a wide variety of applied settings.[15]

The use of IRT to estimate item and ability parameters requires specialized computer software (e.g., LOGIST or BILOG), which at present is not included in any of the major statistical packages commonly used by social scientists (e.g., SPSSx, SAS, BMDP). However, useful information about the discrimination parameter for the item characteristic curve can be obtained with the logistic regression module found in most statistical packages.

The item characteristic curve simply plots the probability of a correct response to a given item as an increasing function of the latent ability of the test taker. In item response theory, the shape of this function is typically hypothesized to be that of the normal ogive or the logistic curve. Figure 1 shows three hypothetical ICCs. Ability is on the x axis, while

[13]The literature on item response theory is considerable. For useful overviews, see Hambleton, Swaminathan, and Rogers (1991); Hambleton and Swaminathan (1985); Thorndike (1982); Allen and Yen (1979); and Lord (1980). Latent trait theory is described by Lazarsfeld and Henry (1968).

[14]Two critical assumptions are (1) unidimensionality of the trait being measured and (2) local independence of the test items (the assumption that for any pair of items on a test, only a test taker's ability affects the likelihood of correctly answering each of them—in the language of causal analysis, that no spurious influences will affect the correlation between the items).

[15]Sample invariance means that an item's performance parameters will hold regardless of the sample to which it is administered (unlike the p value and item-total correlations, which would vary according to the mean and variance of the sample's ability level). The sample invariant quality of IRT measures has led to their application in discovering biased items in standardized tests.

Figure 1. Hypothetical Item Characteristic Curves

Probability of Correct Response

Ability (Standardized)
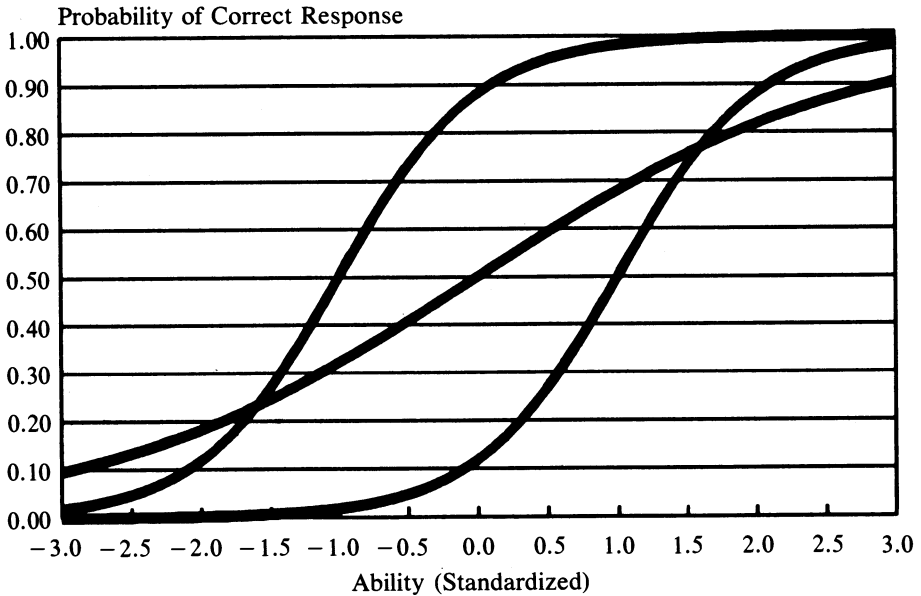
the probability of a correct response is on the y axis. The curve on the left is typical of a relatively easy item that appears to discriminate rather well (the steepness of the curve indicates the extent to which the probability of a correct response increases rapidly relative to changes in ability). The middle curve is a more difficult item but with poorer discrimination. Increases in ability do not bring very dramatic increases in the probability of a correct answer. The curve on the right would have the same discriminating power as the one on the left, but its discrimination occurs at a higher level of ability.

Item characteristic curves can be created by simply plotting the mean p value for all subjects at each of several levels of the total test score. However, this method at best permits only a visual comparison of the steepness of the curves as a way of estimating the discrimination parameters. By using logistic regression to regress a dichotomous test item on the standardized total test score (adjusted by removing the item being tested), estimates of both the discrimination and difficulty parameters can be obtained. The discrimination parameter, which is the coefficient from the logistic regression analysis, is proportional to the slope of the ICC at the point where the predicted p value is .5 (the x axis value where this occurs is the difficulty parameter for the item). A larger coefficient implies

a steeper slope and thus greater discrimination. The statistic is analogous to the item-total correlation in that it reflects (comparatively) the presence of greater or lesser amounts of measurement error. Reliability decreases when there is a large range of ability over which we are unable to predict accurately whether a correct response will be given. The discrimination parameter and the item-total correlation tend to be correlated with one another, but not perfectly so. For many of the items described below, the discrimination parameters for the very easy and very hard ones were quite high.[16]

Table 1 presents five statistics for 39 items from the 1989 national survey.[17] The table is sorted by the item-total correlation. Even with a variety of statistical aids (and, to some extent, *because* of the variety), the selection of specific items remains fairly subjective, guided by the goals of the research and influenced by factors not easily quantified. For example, the desire to maintain respondent morale in the interview may necessitate the inclusion of easier items than the item analysis would dictate.

Our item analysis had several goals: to determine whether certain topics could be measured more effectively than others; to evaluate the performance of different question formats; to identify good questions to use (and bad questions to avoid) on subsequent surveys; and, if possible, to select items to use in a short scale for measuring political knowledge. The data in Table 1 support the following generalizations.

First, no particular topic area appeared easier or harder to measure than the others. The "top 10" items (in terms of item-total correlations) included four civics items, four party and people items, and two issues.

---

[16]The difficulty parameter is related to the p value in nearly linear fashion (except at the extremes) and so for our limited purposes does not provide any additional information beyond that of the conventional item analysis statistics. However, when IRT is used with specialized software to estimate ability levels of subjects, the difficulty parameter is a critical variable (see, e.g., Hambleton and Swaminathan 1985).

One also finds a few other kinds of item performance measures in the literature, particularly with respect to item discrimination. One commonly used measure, which is analogous in some respects to the logistic regression coefficient just described, is the difference in the percentage of high- and low-ability subjects who can correctly answer a given item. We computed a high-low difference index for our items, using the bottom and top quartiles as the basis for discrimination. This index was highly correlated with the item-total correlation ($r = .90$), and so it adds very little to the analyis.

[17]In item analysis, it is important that the criterion scale be a reasonable measure of the underlying construct. Unusually poor items may contaminate or weaken the item analysis. An initial examination of item-total correlations led us to reject three of the original 42 items: right to counsel ($-.01$), abortion rights prior to *Roe v. Wade* (.07), and the percentage of the federal budget spent on Social Security (.10).

Second, there was no penalty in using a particular question format. Open- and closed-ended formats were represented among the best and worst performers. For example, one open-ended item (knowledge about arms control) was at the top in item-total correlation and very high on the discrimination parameter, while another with nearly the same p value (knowledge of the effect of high tariffs on U.S. trade) performed quite poorly. Similarly, closed-ended items with as few as two fixed alternatives, despite their susceptibility to guessing, did as well as open-ended questions. It all depended on the item.

As for specific items, some questions with good face validity were mediocre performers: naming one's governor or U.S. representative, stating one provision of the First or Fifth Amendments, and knowing which side the U.S. government supported in El Salvador were below average on all performance measures. As a general rule, items requiring respondents to state a number or a percentage did poorly, though the unemployment rate was an exception to this.

Finally, the item analysis indicated that a short scale, covering a modest range of topics, is feasible. For example, according to the multiple regression analysis, five items explained over three-quarters of the variance in the full 39-variable measure, and 10 variables accounted for nearly 90% of the variance.

Information Items on the National Election Study Surveys

Building on the analysis above, we turn now to a consideration of knowledge measures found on the NES surveys. These surveys, conducted by the University of Michigan's Center for Political Studies, are the most common source of data for analysts of public opinion and voting behavior. Yet knowledge measures on the NES surveys are rare. The NES's laudable practice of repeating questions in unchanged form over time has, with few exceptions (e.g., party control of Congress or naming House candidates), not extended to the direct measurement of knowledge.

Nevertheless, the structural analyses presented above imply that acceptable measures of political knowledge could be built from the items found on most NES surveys, given the unidimensional quality of the construct.[18] However, the item analysis of our 1989 survey also suggests

[18]We replicated the dimensional analysis (and the construct validation) using two sets of the National Election Study's surveys: the 1984–85 surveys and the 1990–91 surveys. While the mix of items in the NES studies was quite different from our own survey, the evidence nevertheless supports the model of political knowledge as a set of theoretically plausible and highly intercorrelated subdimensions (Delli Carpini and Keeter 1992, 1993).

Table 1. Item Performance Statistics for Questions from the 1989 National Survey of Political Knowledge

	Proportion Correct (p value)	Discrimination Parameter (Logistic Regression Coefficient)	Difficulty Parameter (Logistic Regression Intercept)	Corrected Item-Total Correlation	Step on Which the Item Entered a Multiple Regression (R^2 after Inclusion)
Mean for all items	.48	1.11	−.08	.39	
Mean for top 20 items	.53	1.37	.08	.47	
Arms control knowledge	.47	1.44	−0.17	.54	1 (.34)
Knows "Bill of Rights"	.46	1.42	−0.27	.53	2 (.52)
U.S. supported Contras	.66	1.48	0.95	.53	4 (.72)
Judicial review	.66	1.36	0.88	.51	
Veto override %	.34	1.42	−0.92	.51	8 (.84)
Appoint judges	.58	1.27	0.41	.50	
Party control of House	.68	1.35	1.04	.50	
Name vice president	.74	1.45	1.46	.49	
Party control of Senate	.55	1.15	0.26	.47	7 (.82)
FDR party ID	.63	1.17	0.67	.47	3 (.64)
Truman party ID	.58	1.13	0.42	.46	
Government budget deficit	.78	1.40	1.73	.46	
Name U.S. senators	.25	1.39	−1.52	.45	5 (.76)
U.S. has trade deficit	.82	1.48	2.08	.45	6 (.79)
What is "recession"	.57	1.04	0.35	.44	9 (.86)
New Deal knowledge	.15	1.81	−2.61	.43	

Describe superfund	.12	2.17	-3.28	.43
Percent unemployed	.24	1.27	-1.54	.42
Nixon party ID	.78	1.20	1.60	.42
5th Amendment	.50	0.92	0.02	.40
Communist run for president?	.50	0.85	-0.01	.38
Women's suffrage	.90	1.57	2.99	.37
Name governor	.73	0.92	1.19	.36
Name U.S. representative	.29	0.89	-1.06	.35
1st Amendment	.35	0.82	-0.73	.35
Effect of high tariff on U.S. trade	.52	0.76	0.08	.34
Pledge of allegiance	.75	0.88	1.31	.34
Rehnquist ideology	.30	0.82	-0.96	.33
Who declares war	.34	0.73	-0.74	.31
Education spending	.24	0.79	-1.29	.30
Can your state prohibit abortion?	.72	0.70	1.07	.29
Date of New Deal	.12	1.08	-2.39	.29
Percent black	.12	0.99	-2.28	.28
Percent poor	.18	0.81	-1.68	.28
Size of U.S. budget	.49	0.57	-0.07	.27
U.S. supported whom in El Salvador	.43	0.49	-0.30	.23
Length of presidential term	.96	1.19	3.66	.21
Date of women's suffrage	.10	0.78	-2.49	.20
Defense spending	.28	0.41	-0.97	.18

10 (.88)

that some measures are better than others, and so we shall focus our discussion on the comparative merits of specific items.

Some sophisticated analysis of the NES knowledge items has already been conducted. A module of political knowledge items, proposed and analyzed by Zaller (1986) and Iyengar (1986, 1990), was included on the 1985 NES pilot.[19] Although Iyengar and Zaller differed somewhat in the conclusions they drew, both agreed that the benefits of using domain-specific measures (e.g., knowledge of foreign affairs) were relatively modest, especially for most types of analyses scholars are likely to perform.

A second conclusion from the 1985 pilot study was that the new measures of political knowledge performed little better than the direct and indirect items included on the typical NES survey. The exceptions to this were the new questions that asked respondents to identify the jobs of several political figures. Beginning in 1986, six or seven such items—covering a range of difficulty—have appeared on each survey.

The 1991 NES Pilot

From our perspective, the information items on the 1985 pilot survey had one major limitation: knowledge of institutions and processes was insufficiently probed. To evaluate a broader measure of knowledge, the NES Board placed several civics knowledge items on the 1991 pilot survey: the four best civics items from our 1989 survey (Bill of Rights, judicial review, veto override majority, and judicial nomination), along with two items used on the 1972 NES (length of a senator's term and the number of times a person can be elected president).[20]

In order to conduct an item analysis, the civics items were combined with measures from the 1990 survey to construct a 20-item knowledge index. This index scaled well (alpha = .87) and had good correlations with several criterion variables. Table 2 describes the knowledge variables, while Table 3 presents the performance data for the component items.

Most of the questions did well in the item analysis, as 15 of the 20 had corrected item-total correlations of .45 or higher. Three of the remaining five were difficult identification items whose correlations were low because of their extreme p values; these items had high values on the

---

[19]In recent years, the NES has conducted pilot surveys in odd years to perform experiments and test new questions. The typical design for the pilot surveys involves telephone interviews with a subset of respondents from the previous year's NES sample.

[20]The pilot study was conducted by telephone during the summer of 1991. The civics items were administered to one of three randomly divided subsamples, which had an N of 449 individuals.

Table 2. Description of Items in the 1990–91 NES Surveys

Variable	Variable No. (NES)	% Correct	% Incorrect or Incomplete	% "Don't Know"
People:				
Quayle	v395	84	1	14
Gorbachev	v398	71	14	15
Thatcher	v399	53	29	18
Name one candidate (and his/her party) for U.S. House	v111	23	11	66
Mandela	v400	17	51	32
Foley	v401	12	10	78
Rehnquist	v397	5	19	76
Mitchell	v396	3	12	85
Party:				
Relative ideological location of the two parties	v413; v414	57	25	18
Party with most seats in the House	v402	55	16	29
Relative location of parties on defense spending	v443; v444	52	23	25
Party with most seats in the Senate	v403	47	17	36
Relative location of parties on federal spending	v456; v457	45	26	29
Relative location of parties on aid to blacks	v449; v450	42	30	28
Civics:				
Times a president can be elected	v2852	73	16	11
Whose responsibility is judicial review?	v2849	68	23	9
Whose responsibility is it to nominate federal judges?	v2850	51	32	17
What are the first Ten Amendments called?	v2848	43	6	50
What majority is needed to override a presidential veto?	v2851	37	17	46
How long is a senator's term?	v2853	25	49	26

Source: National Election Study surveys for 1990 and 1991 conducted by the Survey Research Center, University of Michigan. $N = 449$.

discrimination parameter. Items from the three major domains performed comparably. Of the items in the top 15, six were party questions, five were civics questions, and four were politician ID questions. The politician items in the top 15 tended to have higher discrimination coefficients (mean = 1.92) than the civics or party questions (1.45 each).[21]

More specifically, the four civics items from our national survey performed about as well on the NES survey as they had in the 1989 survey. The item-total correlations were similar (a mean of .51 in the 1989 survey, compared with .53 in the NES), as were the discrimination parameters (means of 1.37 and 1.48, respectively).[22]

What about the items traditionally found on the NES surveys? In general, they did very well. The best question in terms of item-total correlation was "House party" (.60); it also had a very high discrimination parameter (1.79), especially for an item susceptible to guessing. The "Senate party" item was good but did not discriminate as well as the House item (a result replicated in the 1989 survey).[23] The party placement questions were generally strong performers, too. The best in terms of both item-total correlation and discrimination was the ideology scale, followed closely by defense spending.

The other item in the index that is commonly found on the NES surveys is naming a House candidate. It was a fairly weak performer, with an item-total correlation of .40 and a discrimination parameter of 1.15. This is not surprising, given the considerable variation in the prominence of House races across the nation, an external factor that undoubtedly introduces its own variance into the measure. We dropped this item in subsequent analyses.

We tested a few additional items commonly found on the NES surveys, though for various reasons we did not include them in the index. In his analyses of presidential elections using NES data, Kessel (1988) measured knowledge with a simple index of the total number of likes and dislikes about the parties and presidential candidates. While this measure rewards the garrulous and penalizes the laconic (and thus also measures a trait other than political knowledge), our analysis suggests that it is nevertheless a valid measure of knowledge—the item-total correlation was .57. Since this measure is found on NES surveys dating back to 1952,

---

[21]Politician feeling thermometer items that included a response of "doesn't recognize" were also tested; their performance was adequate. They would be acceptable components of a knowledge scale if direct identification measures were not available.

[22]The other two civics items (length of a senator's term, and number of times a person can be elected president) did less well and were dropped in subsequent analyses.

[23]It should be noted, though, that these items get some of their strength from each other, since they are so highly intercorrelated.

Table 3. Item Performance Statistics for Questions from the 1990–91 NES Surveys

	Proportion Correct (p value)	Discrimination Parameter (Logistic Regression Coefficient)	Difficulty Parameter (Logistic Regression Intercept)	Corrected Item-Total Correlation	Step on Which the Item Entered a Multiple Regression (r^2 after Inclusion)
Mean for all items	.43			.47	
House party	.55	1.79	.35	.60	1 (.43)
Veto override %	.37	1.67	−.81	.58	3 (.72)
Senate party	.47	1.42	−.18	.54	
Ideological party	.57	1.48	.43	.54	2 (.61)
Defense party	.52	1.43	.14	.53	
Judicial review	.68	1.64	1.18	.52	5 (.82)
Nominate judges	.51	1.36	.08	.52	
Mandela	.17	2.16	−2.70	.52	10 (.93)
Thatcher	.53	1.29	.16	.51	4 (.78)
Gorbachev	.71	1.72	1.43	.51	8 (.90)
Spend party	.45	1.30	−.28	.51	
Black party	.42	1.26	−.42	.49	6 (.85)
Bill of Rights	.43	1.24	−.35	.49	
Quayle	.84	2.52	3.20	.48	
Senator's term[a]	.25	1.32	−1.50	.45	7 (.88)
Name one House candidate[a]	.23	1.15	−1.54	.40	9 (.91)
Foley	.12	1.76	−2.98	.40	
Rehnquist	.05	2.59	−5.14	.33	
Times a president can be elected[a]	.73	.71	1.08	.29	
Mitchell	.03	2.95	−6.43	.28	

[a]Dropped in subsequent analyses

Source: National Election Study surveys for 1990 and 1991 conducted by the Survey Research Center, University of Michigan. $N = 449$.

its value in analyses over time is considerable. A potential limitation, as with several other NES items adapted to the measurement of information, is that the likes-dislikes items are commonly used as dependent variables in substantive analyses; one would not want to use such items on "both sides" of the equation.

A Recommended Five-Item Knowledge Index

The analysis above suggests that researchers using NES data can adequately measure the general concept of political knowledge with available items. However, those collecting their own data may have a need for a dependable knowledge index that is more parsimonious than the large set of NES items just discussed. An important application of item analysis techniques is the derivation of short scales that provide reliable and valid measurement with relatively few individual items.

Using the data in Table 3, we derived a five-item index from the 1990 and 1991 NES questions. The items, and the rationale for their inclusion, are

> *Party control of the House.* A good performer in our 1989 survey and picked first by the stepwise regression in the 1990–91 data, this item has good discriminating power as measured by the logistic regression (fifth overall). It has strong face validity.

> *Veto override percent.* Picked third by the regression, this is the most difficult of the recommended items. Its discriminating power is good and was strong in the 1989 survey.

> *Party ideological location.* Picked second by the regression, this item had the highest discriminating power of the four party placement questions. As another key concept at the heart of contemporary U.S. politics, it has strong face validity.

> *Judicial review.* A relatively easy civics item (68% correct), this question had good discriminating power. It was selected fifth by the regression analysis. The veto item taps familiarity with both Congress and the presidency, while this item ensures that the judiciary is represented.

> *Quayle (identifying the vice president).* This item had high discriminating power according to the logistic regression and was also a strong variable in our 1989 national survey. It is the easiest of the NES items tested, serving to distinguish those who are completely disconnected from politics. And this is one "people" variable that

may remain relatively consistent in its contribution to the scale over time. Despite Dan Quayle's high profile, the percentage of the public able to name him in the 1989 survey, or to identify his position in the 1990 NES survey, was comparable to that for other vice presidents: Alben Barkley in 1952, Richard Nixon in 1953, and Walter Mondale in 1978.

This five-item index scaled well (alpha = .71). Three of the items provide comparability with NES data, while one of the two civics items (judicial review) was added to the 1992 NES general election survey and may be used regularly on future studies.[24] And, as we shall see in the next section on validity assessment, this index was a strong performer in comparison with longer indices.

Assessing the Validity of Knowledge Indices

In its broadest sense, the *validity* of a measure speaks to the kinds of generalizations that can be made from it (Thorndike 1982, 184). Since the underlying trait of interest cannot be measured directly, behaviors that can be observed are taken to stand for the concept. *Which* behaviors are taken is the issue of *content validity,* which was addressed earlier. Once we have a measure that we believe is representative of various behaviors associated with a trait, we can evaluate its performance in terms of the theories that make use of the trait. This process assesses *construct validity.* Paraphrasing Nunnally (1978, 98), a key question for construct validation is whether studies using the measure produce results that are predictable from highly accepted theoretical hypotheses concerning the construct.

To illustrate the process of construct validation, as well as to assess the performance of different indices of political knowledge, we created six different active scales with the 1990–91 NES data: a 17-item "deluxe" model (alpha = .87); a 13-item scale that consists of all the items that are regularly available in the NES election year surveys (alpha = .83); three "domain-specific" scales that measure knowledge about civics, people, and parties, respectively (alphas = .68, .69, .79); and the recommended five-item short-scale (alpha = .71).

Political knowledge is expected to be related to various political behaviors such as participation, efficacy, and opinionation. Thus, we should see positive correlations between measures of knowledge and measures

---

[24]An additional civics item—appointing judges—is also included on the 1992 NES survey.

of these behaviors. Table 4 shows the correlations between eight criterion variables and the NES indices.[25]

Not surprisingly, all but one of the criterion variables were significantly correlated with the scales in ways consistent with both theory and prior research. While this finding is encouraging, substantively and statistically significant validity coefficients are only a *necessary* and not a *sufficient* condition to establish the validity of the measures. More important is the *relative* pattern of these relationships across competing measures. Three main conclusions can be drawn from a comparison of the correlations presented in Table 4.[26] First, while the party index performs the best of the three domain-based indices, the correlations across the three are nearly identical. That the party index would do well is not terribly surprising, given that the criterion variables tend to measure engagement in contemporary partisan elections in the United States. More surprising—and a testament to the fundamentally general nature of political knowledge—is that a four-item civics knowledge index does almost as well on most measures.[27]

Second, the scale based on the regularly available NES items performs about as well as the "deluxe" index, which includes the civics items. The civics items have many virtues, including their relative timelessness when compared with the party and people items, which measure

---

[25] A description of the criterion variables is available from the authors. One of the criterion variables is the five-category interviewer rating of respondent's information level. We believe that any index of political knowledge built with the NES items should also include this variable. As its high correlation with our knowledge indices attests, it appears to be a highly valid measure. Zaller has shown that the interviewer rating is relatively uncontaminated by "status bias" in favor of higher status respondents (1985, 4–6) and that the measure has good discriminating power, especially at lower levels of respondent knowledge (1986, 17–19).

Despite its power in the NES analyses, the interviewer rating would not be a sufficient measure of political knowledge for most surveys. First, the interviewers undoubtedly base some of their judgment on the respondents' answers to the knowledge questions. Thus, without the direct questions, the interviewers' ratings might be less valid. Second, the reliability and validity of interviewer ratings is likely to be much higher in the NES surveys than in many others. NES interviewers are well trained and usually very experienced, and the lengthy, face-to-face interviews provide an ideal setting for assessing a respondent's level of political engagement and sophistication. Interviewer ratings based on shorter, less comprehensive telephone surveys, are apt to be less valid and reliable.

[26] Similar analyses were conducted with the 1989 data and yielded comparable results.

[27] The party index is the best of the three but is also the lengthiest to administer. The six component items require 10 questions to construct because of the placement items (14 if one counts the filter items). Because knowledge of party politics emerged as a somewhat independent factor in our dimensional analyses, depending *exclusively* on such items might bias one's further interpretation in some small but potentially significant ways.

Table 4. Correlation of NES Indices with Criterion Variables

	Civic Index (4 items)	Party Index (6 items)	People Index (7 items)	13 NES Items (Not Including Civic Items)	17 NES and Civics Items	Best 5 Items
Civic index	1.00	.58**	.63**	.67**	.83**	.82**
Party index	.58**	1.00	.60**	.93**	.89**	.80**
People index	.65**	.61**	1.00	.85**	.85**	.72**
NES items (not incl. civics)	.68**	.92**	.85**	1.00	.97**	.85**
NES items and civics items	.83**	.88**	.85**	.97**	1.00	.91**
Interviewer rating of information level	.57**	.59**	.59**	.66**	.68**	.62**
Efficacy	.27**	.37**	.36**	.41**	.39**	.34**
Participation	.39**	.42**	.37**	.44**	.46**	.43**
Ideological stability, 1990–91	.23**	.26**	.21**	.27**	.28**	.23**
Defense spending stability, 1990–91	.21**	.25**	.22**	.26**	.27**	.23**
Racial attitude stability, 1990–91	.21**	.18**	.15**	.19**	.21**	.16*
Partisan stability, 1990–91	.02	.04	.03	.04	.04	.03
Opinionation	.30**	.39**	.31**	.40**	.40**	.38**

Civics index:
Veto override %
Judicial review
Nominate judges
Bill of Rights

Party index:
House party
Senate party
Ideological party
Defense party
Spend party
Black party

People index:
Mandela
Thatcher
Gorbachev
Quayle
Foley
Rehnquist
Mitchell

Best 5 index:
House party
Veto override %
Judicial review
Ideological party
Quayle

$*p < .05; **p < .01.$

Source: National Election Study surveys for 1990 and 1991 conducted by the Survey Research Center, University of Michigan. $N = 449$.

more ephemeral phenomena. However, this analysis demonstrates that the regular NES items can be used to create a perfectly acceptable measure of political knowledge at a particular time.

Third, the recommended five-item index performs well relative to the two longer scales. Its coefficient alpha was a respectable .71, and correlations with the interviewer rating (.62), participation (.43), opinionation (.38), efficacy (.34), stability of ideology (.23), and stability of defense spending attitudes (.23) were all close to those of the longer scales, despite its being much shorter. This index should provide reasonably comparable discriminating power over time. While we recommend this five-item index, the more general point from this analysis is that very short scales composed of carefully chosen items can measure political knowledge with an acceptable level of reliability and validity.

Using Focus Groups to Validate a Survey-Based Measure of Knowledge

Finally, we address a larger question of validity. Survey measures of political knowledge correlate reasonably well with other survey measures of political behavior. But such measures of validation are inherently limited in that they are all derived from an individual's performance in the same highly artificial setting—the survey interview. It is hard to know, based on the survey data alone, how individuals use their factual knowledge about politics, or more generally, how politically competent they would appear in a more extensive examination of their political beliefs and attitudes. This issue is sometimes referred to as *predictive validity* (Spector 1992, 48).

To address this question, we conducted an exploratory study consisting of four focus groups with individuals who had responded to one of our local telephone surveys (see the appendix for a description). The survey included a four-item knowledge test on national politics. In all, 21 individuals took part in the groups, each of which met for about two hours for a discussion of opinions on various political issues. The transcribed interviews were coded to ascertain each participant's use of factual information.[28] Of all the social and political variables available to us (including interest in politics, media use, education, income, etc.), the

---

[28]Our rules for coding the use of political facts were similar to those employed in the analysis of the Bay Area Survey depth interviews, utilized by Neuman (1981). Fact totals for each individual were standardized according to the share of time available to each participant in his or her session (based on the number of participants). Details of this part of the study are available from the authors.

national knowledge survey index was the best predictor of use of facts in the focus groups (simple $r = .51$; beta $= .48$; R-squared for the regression $= .51$). Considering the inherent limitations in this method of criterion validation, the results provide reassurance that even a short survey-based measure of political knowledge can assay meaningful variations in the cognitive political sophistication of survey respondents.[29]

Concluding Comments: Measuring Political Knowledge

Valid measurement is a cornerstone of successful scientific inquiry. But for a variety of reasons, measurement issues in the social sciences often receive inadequate attention from researchers. Because of the scarcity of resources for primary data collection, social research is increasingly based on secondary analysis of existing data. As a consequence, scholars must "make do" with the available measures and thus may not feel that extensive attention to measurement issues is worthwhile. Where primary data collection is possible, the logistics often demand the lion's share of researchers' time in the early stages of a project. More generally, in doing research we all want to "cut to the chase" and look for the substantive findings in our data.

A chief goal of this paper has been to appeal for greater attention to measurement issues, even when using existing data. We have illustrated the use of various techniques—some old, some relatively new—for the evaluation of individual measures and their combination into scales. While considerable attention was paid to the mechanics of the process, we hope that the importance of the researcher's theoretically informed judgment was also communicated. Measurement is often as much art as science.

Our discussion of scale development centered more specifically on an important and growing area of the study of political psychology and behavior: political knowledge. Data we examined indicate that political knowledge is a relatively unidimensional concept, that a citizen's level of factual knowledge can be gauged with a short series of survey questions, and that items routinely included on the NES surveys can be used to create a good scale of political knowledge—provided that the proper items are selected. We assessed the validity of several knowledge scales and offered a simple five-item scale for use on surveys. The strong performance of short indices, in conjunction with growing evidence of the

---

[29]We are unaware of other efforts to use focus groups in this fashion for scale validation. However, we found them to be valuable in this process and encourage other researchers to experiment with them as we did.

importance of political information in understanding the dynamics of public opinion, should encourage researchers to include knowledge measures as a standard feature of political surveys.

Manuscript submitted 10 June 1992
Final manuscript received 14 January 1993

APPENDIX

Description of the 1989 National Survey

The survey employed computer-assisted telephone interviewing and a commercially prepared random sample. Respondents within households were selected randomly by the CATI software following a household enumeration. Interviews averaged 23 minutes in length. The sample was weighted on gender, race, and education according to 1986 population estimates of the U.S. Census Bureau. The simple response rate was 38%, while the CASRO rate, which is adjusted for "no answer" and "busy" outcomes, was 36%. Despite the low response rate, the sample appears valid on a number of criteria. Compared with the samples for recent National Election Studies and General Social Surveys, our sample was nearly identical in the distribution of partisanship and very similar in most reported levels of political activity, including turning out to vote, reading a daily newspaper, watching network television news, and discussing politics with family and friends. Knowledge questions available for comparison with the 1991 NES and 1987 GSS also indicate no serious bias. For a further discussion of this survey, see Delli Carpini and Keeter (1991).

Description of the 1991 Local Survey Used to Recruit the Focus Groups

The screening survey interviewed 1,208 residents of the Richmond, Virginia, metropolitan area during March and April 1991. The focus group participants were recruited from the 329 respondents who resided in Chesterfield County, one of the three major jurisdictions of the metro area. The CASRO response rate for the survey was 72%. A commercial random sample was used.

Question Wording for a Recommended Five-Item Knowledge Index

Recommended introduction: Last, here are a few questions about the government in Washington. Many people don't know the answers to these questions, so if there are some you don't know just tell me and we'll go on.

    1. Do you happen to know what job or political office is now held by (insert name of current vice president)? (*Original wording in NES:* Now we have a set of questions concerning various public figures. We want to see how much information about them gets out to the public from television, newspapers and the like. The first name is Dan Quayle: what job or political office does he now hold?

    2. Whose responsibility is it to determine if a law is constitutional or not . . . is it the president, the Congress, or the Supreme Court?

    3. How much of a majority is required for the U.S. Senate and House to override a presidential veto?

    4. Do you happen to know which party had the most members in the House of Representatives in Washington before the election this/last month?

    5. Would you say that one of the parties is more conservative than the other at the

national level? Which party is more conservative? (In the item analysis, the party ideology item is constructed from the respondent's placement of the parties on an ideology scale. However, an item based on the direct question of which party is more conservative works as well and is much easier to administer.)

REFERENCES

Allen, Mary J., and Wendy M. Yen. 1979. *Introduction to Measurement Theory*. Monterey, CA: Wadsworth.

Barber, James David. 1969. *Citizen Politics*. Chicago: Markham.

Bennett, Stephen Earl. 1990. "The Dimensions of Americans' Political Information." Presented at the annual meeting of the American Political Science Association, San Francisco.

Berelson, Bernard R., Paul F. Lazarsfeld, and William N. McPhee. 1954. *Voting: A Study of Opinion Formation in a Presidential Campaign*. Chicago: Univeristy of Chicago Press.

Crocker, Linda, and James Algina. 1986. *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston.

Delli Carpini, Michael X., and Scott Keeter. 1991. "Stability and Change in the U.S. Public's Knowledge of Politics." *Public Opinion Quarterly* 55:583–612.

———. 1992. "An Analysis of Information Items on the 1990 and 1991 NES Surveys." Report to the Board of Overseers for the National Election Studies, Ann Arbor, MI.

———. 1993. "Political Knowledge and the Democratic Citizen." Unpublished manuscript.

Fiske, Susan T., Richard R. Lau, and Richard A. Smith. 1990. "On the Varieties and Utilities of Political Expertise." *Social Cognition* 8:31–48.

Hambleton, Ronald K., and Hariharan Swaminathan. 1985. *Item Response Theory: Principles and Applications*. Boston: Kluwer-Nijhoff.

Hambleton, Ronald K., Hariharan Swaminathan, and H. Jane Rogers. 1991. *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.

Iyengar, Shanto. 1986. "Whither Political Information?" Report to the Board of Overseers and Pilot Study Committee, National Election Studies.

———. 1990. "Shortcuts to Political Knowledge: The Role of Selective Attention and Accessibility." In *Information and Democratic Processes*, ed. John A. Ferejohn and James H. Kuklinski. Chicago: University of Illinois Press.

Iyengar, Shanto, and Donald R. Kinder. 1987. *News that Matters: Television and American Opinion*. Chicago: University of Chicago Press.

Kessel, John H. 1988. *Presidential Campaign Politics: Coalition Strategies and Citizen Response*. 3d ed. Chicago: Dorsey Press.

Krosnick, Jon A. 1990. "Lessons Learned: A Review and Integration of Our Findings." *Social Cognition* 8:154–58.

Krosnick, Jon A., and Michael A. Milburn. 1990. "Psychological Determinants of Political Opinionation." *Social Cognition* 8:49–72.

Lau, Richard R., and Ralph Erber. 1986. "Political Sophistication: An Information Processing Perspective." In *Mass Media and Political Thought*, ed. Sidney Kraus and Richard M. Perloff. Beverly Hills: Sage.

Lazarsfeld, Paul F., and Neil W. Henry. 1968. *Latent Structure Analysis*. New York: Houghton Mifflin.

Lewis-Beck, Michael S. 1978. "Stepwise Regression: A Caution." *Political Methodology* 5:213–40.

Lodge, Milton, Kathleen M. McGraw, and Patrick Stroh. 1989. "An Impression-Driven Model of Candidate Evaluation." *American Political Science Review* 83:399–419.

Lord, Frederick M. 1980. *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum.

Luskin, Robert C. 1987. "Measuring Political Sophistication." *American Journal of Political Science* 31:856–99.

McGraw, Kathleen, and Neil Pinney. 1990. "The Effects of General and Domain-Specific Expertise on Political Memory and Judgment." *Social Cognition* 8:9–30.

Neuman, W. Russell. 1981. "Differentiation and Integration: Two Dimensions of Political Thinking." *American Journal of Sociology* 86:1236–68.

———. 1986. *The Paradox of Mass Politics: Knowledge and Opinion in the American Electorate*. Cambridge: Harvard University Press.

Nunnally, Jum C. 1978. *Psychometric Theory*. 2d ed. New York: McGraw-Hill.

Owen, Diana, and Marianne Stewart. 1987. "Explaining Political Knowledge: Problems of Conceptualization and Measurement." Presented at the annual meeting of the Southern Political Science Association, Charlotte, NC.

Price, Vincent, and John Zaller. 1990. "Measuring Individual Differences in Likelihood of News Reception." Presented at the annual meeting of the American Political Science Association, San Francisco.

Smith, Eric R. A. N. 1989. *The Unchanging American Voter*. Berkeley: University of California Press.

Spector, Paul E. 1992. *Summated Rating Scale Construction: An Introduction*. Newbury Park, CA: Sage.

Thorndike, Robert L. 1982. *Applied Psychometrics*. New York: Houghton Mifflin.

Yankelovich, Daniel. 1991. *Coming to Public Judgment: Making Democracy Work in a Complex World*. Syracuse: Syracuse University Press.

Zaller, John. 1985. "Pretesting Information Items on the 1986 NES Pilot Survey." Memorandum to the Board of Overseers for the National Election Studies.

———. 1986. "Analysis of Information Items in the 1985 NES Pilot Study." Report to the Board of Overseers for the National Election Studies.

———. 1990. "Political Awareness, Elite Opinion Leadership, and the Mass Survey Response." *Social Cognition* 8:125–53.

———. 1991. "Information, Values, and Opinion." *American Political Science Review* 85:1215–37.

———. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.

Zeller, Richard A., and Edward G. Carmines. 1980. *Measurement in the Social Sciences: The Link Between Theory and Data*. New York: Cambridge University Press.