Asked and Answered: Knowledge Levels When We Won't Take "Don't Know" for an Answer

Jeffery J. Mondak
Belinda Creel Davis

Department of Political Science
Florida State University

Abstract. A pivotal claim in research on American political behavior is that the typical citizen knows very little about politics. Public opinion surveys provide a considerable body of evidence in support of this position. However, survey protocols with respect to factual questions about politics violate established norms in educational testing in that "don't know" answers are encouraged rather than discouraged. Because encouraging "don't know" responses potentially confounds efforts to identify substantive understanding, this practice may lead to the systematic understatement of political knowledge. We explore this possibility with data drawn from two split-ballot tests: one conducted as part of a survey in the Tallahassee, Florida metropolitan area and one conducted as part of the 1998 NES Pilot. Results reveal that the mean level of political knowledge increases by approximately 15 percent when knowledge questions are asked in accordance with accepted practices in educational testing.

Political scientists debate so many issues with such great intensity that points of agreement seem oddly conspicuous. One question on which near-universal consensus exists concerns how much Americans know about politics. The answer, of course, is not much at all. It is widely accepted that relatively few Americans can be considered well-informed about politics and government. In stark contrast with this sophisticated minority, most citizens, analysts agree, know little—or very little—about these subjects. Philip Converse captured this sense of scholarly consensus well, writing that (1990, 372) "the two simplest truths I know about the distribution of political information in modern electorates are that the mean is low and the variance high." Writing in the same volume, John Ferejohn stated matters in even stronger terms: "Nothing strikes the student of public opinion and democracy more forcefully than the paucity of information most people possess about politics" (1990, 3). Likewise, Russell Neuman noted that "the available data show overwhelmingly that even the basic facts of political history, the fundamental structure of political institutions, and current political figures and events escape the cognizance of the great majority of the electorate" (1986, 15).

Widespread political ignorance and representative democracy make a poor match. As a result, political scientists have devoted considerable effort to exploring the implications of an uninformed citizenry. Indeed, it is hardly a stretch to suggest that reference to low levels of political information in the mass public constitutes one of the most popular starting points for studies of political behavior. Much of this research has sought to determine whether Americans function as capable citizens, both individually and in the aggregate, in spite of the fact that information levels are so low. Prominent works in this extremely rich research tradition include those by Delli Carpini and Keeter (1996), Page and Shapiro (1992), Popkin (1994), Sniderman, Brody and Tetlock (1991), and Zaller (1992).

No matter how stable a structure may seem, it is wise to inspect the foundation from time to time to ensure that no faults have developed. This same advice holds for enduring scholarly consensus. With this in mind, we have reassessed the distribution of political information. Our results lead us to challenge the conventional wisdom described above. To be clear about our purpose, we will make no absolute

claims regarding either levels of political knowledge or variance in those levels. We will not argue, for instance, that the distribution of political information is such that the mean is high and the variance low. Instead, we will cast the evidence in relative terms. We contend that political scientists have consistently and systematically overstated the basic case regarding political information and the American citizen. Specifically, we will show that knowledge levels are about 10 to 15 percent higher, and variance about 10 to 15 percent lower, than usually reported in empirical research.

It has become increasingly common for researchers to measure political knowledge by asking survey respondents a brief battery of factual questions about politics. We are concerned with the way these questions are worded. Political scientists typically design knowledge batteries in a manner inconsistent with established procedures in item design, procedures that were developed over the course of decades of exhaustive analysis in the field of educational testing. The consequence in terms of measurement is that violation of standard testing practices results in a reduction in the validity of knowledge scales. The consequence in terms of substantive results is that knowledge levels are understated, and variance in knowledge is overstated. In demonstrating these points, our analysis will proceed in two parts. First, we will discuss the approach to measuring political knowledge that is used on most opinion surveys, and we will explain how this approach conflicts with standards in educational testing. Second, we will analyze data from two split-ballot surveys on which each respondent was asked knowledge items using one of two formats: one designed to resemble common practice in political science, and the other designed to incorporate the accumulated wisdom of experts on test construction.

Approaches to the Measurement of Political Knowledge

All students of public opinion recognize that even seemingly minor changes in question wording, question order, interviewing techniques, and so on can yield sizable swings in substantive findings. We will demonstrate such an effect below when we compare political knowledge data gathered using two somewhat different survey protocols. But where different procedures produce different results, on what basis should we decide which procedure to employ? We believe that validity should be the deciding

factor. Given a choice between two measurement procedures, analysts should opt for the one that best captures the construct in question and best avoids systematic error.

The usual approach to measuring levels of political knowledge entails asking survey respondents a series of questions concerning basic factual matters such as the identity of the vice president and which party controls the House of Representatives. Common formats include the two-category multiple-choice item, the three-category multiple-choice item, and the open-ended identification item. A respondent can answer any knowledge item in one of three ways: by stating the correct response, by stating an incorrect response, or by answering "I don't know" (DK). Knowledge scales usually are constructed by counting the number of correct answers given by each respondent; consequently, it follows that incorrect responses and DKs are assumed to indicate the same thing (i.e., the absence of knowledge).

Regardless of what procedure is employed, it is impossible for any item designed to measure political knowledge to achieve perfect validity. Hence, although validity is an appropriate criterion for selection between two measures, it would be futile to hold out for the perfect indicator. The problem stems from the complexity of knowledge as a construct. When considering levels of knowledge, four behavioral states exist: 1) fully informed (i.e., the respondent truly does know the answer to our question); 2) partially informed (the respondent either possesses an incomplete understanding, or the respondent can rule out an incorrect choice option on a multiple-choice item); 3) misinformed (the respondent believes he or she knows the correct answer, but is mistaken); and 4) uninformed (the respondent holds no knowledge pertinent to our question). The difficulty in terms of measurement is that our three choice options—correct, incorrect, and DK—simply do not line up precisely with the four underlying behavioral states.

The relationship between survey responses and actual behavioral states is summarized in table 1. Only two of twelve cells are empty.[1] First, if a person answers an item correctly, we can infer that this individual is not misinformed. Second, if a person answers incorrectly, we can infer that this individual is

not fully informed. No other relationships can be ruled out. All instructors recognize that a correct answer does not necessarily mean that the test-taker "knows" the subject matter in question. On surveys, a correct answer can be recorded when the respondent is fully informed, partially informed (the correct educated guess), and even uninformed (the correct blind guess). Likewise, an incorrect answer can be recorded when the respondent is misinformed, partially informed (the incorrect educated guess), and uninformed (the incorrect blind guess). Finally, respondents may answer DK regardless of how much or how little they actually know.

The problem with the loose fit between survey response options and knowledge as a construct is that any single knowledge item will tend to be unreliable. As one example, if one uninformed respondent guesses correctly and another guesses incorrectly, their scores will differ because of chance, not because one is more knowledgeable than the other. But this does not imply that it is hopeless to attempt to measure knowledge. The problem described here can, to a large extent, be overcome through use of data from several items in a summary scale. Indeed, this is one reason why teachers include more than just a handful of items on their multiple-choice examinations. On a 100-item test, for example, meaningful results will be obtained. An uninformed student who guesses on each question will answer some items correctly, but the uninformed student will receive a lower score than the partially-informed student, and the partially-informed student will fare worse than the fully-informed student.

Although multiple-item scales can provide good measures of political knowledge, common survey procedures in political science carry an additional threat to validity. Surveys used to measure political knowledge diverge from norms in educational testing on the matter of "don't know" responses, and it is on this point that we are most concerned. Few surveys discourage DKs, and, in many cases, DKs are invited. Delli Carpini and Keeter explicitly recommend that DKs be encouraged. They suggest that knowledge batteries begin with the following introduction: "Last, here are a few questions about government in Washington. Many people don't know the answers to these questions, so if there are some

[1] Two cells are empty only if we assume no recording error on the part of survey interviewers.

you don't know just tell me and we'll go on" (1996, 305). Individual items also are often worded so as to encourage DKs, such as by prefacing questions with the phrase "Do you happen to know..." When respondents do answer "don't know" on knowledge items, survey interviewers routinely are instructed not to prompt the respondents for substantive answers. All of this effort to encourage DKs apparently achieves the desired end. DK rates on knowledge items generally exceed DK rates on attitudinal items, and most knowledge questions generate more DKs than wrong answers. When encouraged, people are willing to admit that they do not know the answers to questions about political knowledge.

One apparent benefit of encouraging DKs is an increase in the reliability of knowledge scales (Delli Carpini and Keeter 1993). If all respondents who answer DK were instead to offer random guesses, the infusion of unsystematic variance would lower inter-item correlations, thereby decreasing reliability. Delli Carpini and Keeter show that a five-item scale can attain acceptable reliability; more items might be needed if DKs were not encouraged. A second pragmatic benefit of encouraging DKs is that this practice may make the survey interview less disconcerting for poorly-informed respondents. However, Delli Carpini and Keeter (1996, 295) report that “fears of knowledge testing’s damage to rapport are exaggerated,” and that they detected no evidence that levels of cooperation declined because knowledge questions were asked on their surveys. Delli Carpini and Keeter asked their respondents 39 knowledge questions, and thus their results suggest that it is highly unlikely that more abbreviated knowledge batteries would be disconcerting for respondents. New evidence supports this claim. Below, we present split-ballot data from the 1998 NES Pilot study, where DKs were encouraged for half of respondents and discouraged for the other half. NES interviewers answer several questions regarding the demeanor and reactions of respondents. The knowledge split ballot produced no effects on these assessments; most importantly, discouraging DKs did not increase respondents’ levels of confusion or agitation, or expressions of doubt or embarrassment regarding any lack of knowledge.

Whatever the merits of encouraging DKs for reliability and interview rapport, it should be clear that neither of these factors stems from concern with validity. That is, the norm of encouraging DKs on

knowledge batteries may be in place for reasons unrelated to whether doing so helps or hurts in the effort to measure knowledge accurately. One counterargument to this claim is that encouraging DKs may, in fact, increase validity, because correct answers now do not reflect the lucky guesses of uninformed respondents. However, this counterargument implicitly introduces two implausible assumptions: that encouraging DKs discourages substantive responses *only* from uninformed respondents, and that *all* uninformed respondents willingly refrain from guessing.

A second counterargument is that political scientists perhaps should be more concerned with *accessible* knowledge (as opposed to knowledge stored in long-term memory), and encouraging DKs may facilitate valid measurement of accessible knowledge. We see several problems with this position. First, Delli Carpini and Keeter (1996) argue persuasively for the importance of knowledge stored in long-term memory as an analytical construct, and we share their view. In contrast, we are aware of no research that establishes the validity of current survey protocol as a means to measure accessible knowledge. Second, encouraging DKs does not ensure that only accessible knowledge is measured. Some respondents do search their memories when answering current knowledge batteries; that is, some respondents try very hard to do their best (Delli Carpini and Keeter 1996). This means that encouraging DKs possibly causes us to measure different things for different people: accessible knowledge for those who do not think carefully about knowledge questions; stored knowledge for those who do. Hence, current survey protocol does not permit valid inference regarding levels of accessible knowledge. Third, irrespective of whether encouraging DKs facilitates the measurement of accessible knowledge, this practice brings severe threats to validity on other grounds. Encouraging DKs invites a guessing response set in which respondents' scores vary as a systematic function of personality traits. Fourth, available empirical evidence provides no support for the claim that current procedures measure accessible, as opposed to stored, knowledge. On the NES split ballot and on a similar survey we conducted locally, half of respondents were discouraged from answering DK. If this change in procedure prompted respondents to search their long-term memories rather than merely calling up readily accessible information, then the average length of

interviews should differ for the two halves of each split-ballot test. No such differences exist. Interview length does not vary by a significant margin as a function of the split ballot on either survey; local surveys were slightly longer when DKs were encouraged (a difference of 12 seconds for 23-minute interviews), whereas NES surveys were slightly shorter.

Drawing on literature from the field of educational testing, we have become persuaded that the validity of political knowledge scales suffers as a consequence of efforts to promote DK responses. For decades, researchers in educational testing have agreed that DKs should be discouraged, not encouraged, on tests of factual knowledge. The logic motivating this advice is that when DKs are encouraged, knowledge batteries systematically tap two distinct constructs: knowledge, and the propensity to guess. The propensity to guess varies as a function of respondents' personality traits. Consequently, when DKs are encouraged, the resulting "knowledge" data may be contaminated. The earliest warning of this possibility was issued over 60 years ago by Wiley and Trimble (1936), who argued that objective tests could be used as a means to measure personality. Today, there is overwhelming agreement that DKs should be discouraged (e.g., Brown 1970; Cronbach 1942, 1946; Cunningham 1986; Ebel and Frisbie 1986: Mehrens and Lehmann 1984; Nunnally 1972). Researchers have found that efforts to encourage DKs invite variance in knowledge scores driven by respondents' levels of self-confidence (Casey, Nutall and Pezaris 1997; Gritten and Johnson 1941; Hirschfeld, Moore and Brown 1995; Sheriffs and Boomer 1954; Stanley and Hopkins 1972), propensity to take risks (Ben-Shakhar and Sinai 1991; Cronbach 1946; Sheriffs and Boomer 1954; Slakter 1969; Swineford 1941), and competitiveness (Hirschfeld, Moore and Brown 1995).

Consider a scenario in which two respondents are each 90 percent certain that the Republicans hold a majority in the U.S. House. When asked, one of these respondents offers a substantive reply, and is awarded credit for giving a correct answer; the other respondent answers DK, and is thus given no credit. The respondents are equally knowledgeable, but they receive different scores wholly because one offered a substantive answer and the other did not. Because of scenarios such as this, researchers in

educational testing argue that knowledge is systematically understated for some test-takers when DKs are encouraged (e.g., Aiken 1988; Sheriffs and Boomer 1954; Stanley and Hopkins 1972). In a series of studies conducted in the 1960s, Slakter (1968a, 1968b, 1969) found that the protocol of encouraging all test-takers to answer every item best ensures that all test-takers will capitalize on their partial knowledge.

The differential propensity to guess is a *response set*, or *response style*. Cronbach (1946, 491) explained that "response sets are defined as any tendency causing a person consistently to make different responses to test items than he would have made had the same content been presented in a different form." Cronbach further explained that (1946, 491) "because they permit persons with equal knowledge, identical attitudes, or equal amounts of a personality trait to receive different scores, response sets always reduce logical validity." This is the scenario we face with conventional measures of political knowledge.

Elsewhere, we have presented empirical evidence regarding the risks described here (Mondak 1999). Several findings warrant emphasis. First, DKs and incorrect answers performed dissimilarly as both independent variables and dependent variables in tests using data from the 1992, 1994, and 1996 National Election Studies. Hence, the inference that these response categories represent an identical state (i.e., the absence of information) is invalid, and the common practice of scaling these categories identically is inappropriate. Second, analysis of these same data reveals that respondents who are uninterested in the survey are inclined to answer DK with great frequency, implying that low knowledge scores may partly reflect apathy rather than ignorance. Third, analysis of data from both the 1958 NES and a recent survey of undergraduates shows that DKs decrease, and thus correct answers increase, for respondents with the highest levels of self-confidence and propensity to take risks.

The accepted practice in educational testing is that all students be encouraged to answer every test item, and that no explicit DK option be offered. For those few students who do elect to omit some answers, Kline (1986) recommends that these DKs be assigned randomly to the available substantive response categories as a means to mimic blind guessing, thus ensuring that these students are not punished for nonresponse. We advocate implementation of comparable procedures on survey batteries

designed to measure political knowledge. This recommendation is based both on our review of the testing literature and on our empirical assessment of extant knowledge batteries (Mondak 1999).

The approach we have recommended can be expected to bring several consequences. First, validity will increase. Current knowledge scales systematically measure two factors, knowledge and the propensity to guess, and one unsystematic factor, chance. Propensity to guess is eliminated if all respondents answer every item, meaning that the sole remaining systematic source of variance is knowledge. Under the approach we recommend, the "don't know" column in table 1 will vanish. This means that fully-informed respondents who would have answered DK now will answer correctly, the misinformed will answer incorrectly, and the partially informed and the uninformed will get some right and some wrong, with success determined by the level of partial information and by chance.

A second consequence is that reliability will decrease. The approach we advocate entails trading a systematic source of variance (propensity to guess) for an unsystematic source of variance (chance). This trade-off has long been recognized in educational testing. For instance, Cronbach (1946), a pivotal figure in analysis of reliability and validity, argued that the increase in random error brought by encouraging test-takers to answer all items is well worth the cost given the benefits that will be accrued in terms of heightened validity. In advocating that DKs be encouraged, Delli Carpini and Keeter (1993) prioritized reliability over validity. We disagree with that decision. It does us no good to have a reliable scale if that scale suffers from systematic contamination. We should not desire reliability that is artifactual. As Cronbach explained (1950, 22) "there is no merit in enhancing test reliability unless validity is enhanced at least proportionately." Thorndike and Hagen concurred, writing that validity "is the crucial test of a measurement procedure. Reliability is important only as a necessary condition for a measure to have validity" (1969, 189).

The third probable consequence, and the one we are most interested in here, is that aggregate knowledge scores will increase, and variance will decrease. If, given the opportunity, a respondent would have answered "don't know," there are two ways this person may come to answer correctly when DKs are

discouraged. First, the respondent may be fully or partially informed. In short, the respondent *does* know. Again, this is what Slakter found in his series of studies in the 1960s: students who could have answered correctly were deterred from doing so when DK was available as a choice option, and thus resulting data systematically understated students' knowledge levels. Second, the uninformed respondent may answer a multiple-choice item correctly purely by chance. Multiple-choice items commonly included on public opinion surveys offer only two or three choice options, meaning that converting DKs to substantive answers is virtually assured of elevating knowledge scores.

Readers understandably might find this last point to be troubling. Is it not problematic that the measurement protocol we recommend will encourage blind guessing? We have several comments regarding this concern. First, guessing undoubtedly already occurs on political knowledge questions. It is impossible to eliminate all guessing because there is no way to force uninformed respondents to answer DK.[2] Hence, the issue ultimately is not whether guessing is to be permitted, but instead whether it is preferable for only some respondents to view guessing as permissible. As explained above, it is our view that all respondents should play on the same field. Knowledge scores should not be allowed to differ from one respondent to the next as a systematic function of willingness to guess. On the NES, for instance, consider a case in which some uninformed respondents guess on each item while other uninformed respondents answer "don't know" on each question: those who guess will, on average, answer at least two questions correctly, whereas those who do not guess will answer zero correctly.[3] Hence, observed "knowledge" will vary as a systematic function of propensity to guess—a classic example of a response set. Currently, we have no means to determine which respondents who offer substantive answers are merely guessing, and we have no means to determine whether DKs are masking partial, or even full, information. If all respondents are encouraged to answer every item, then differential

[2] Open-ended questions are relatively immune to blind guessing, particularly when compared with multiple-choice formats, but open-ended questions may exacerbate the personality-based contamination discussed above.

[3] The1992 and 1994 renditions of the NES included three knowledge items with two choice options and two items

propensities to guess will no longer skew results, and information will no longer go undetected. Again, the bottom line is validity.

Second, assuming that aggregate knowledge scores increase under the measurement protocol advocated here, it is possible to estimate what portion of that increase is attributable to guessing and what portion is attributable to truly informed response. Some of the survey questions analyzed below use an open-ended identification format. These questions are not susceptible to blind guessing. Hence, an increase in scores on these items would provide evidence that encouraging DKs causes some knowledgeable respondents to be identified as uninformed. Likewise, on multiple-choice items it can be determined whether any increase in scores exceeds what would be expected on the basis of blind guessing alone. Increased guessing is a minor side effect if our treatment brings the advantage of correcting systematic understatement of knowledge levels.

Guessing constitutes a source of measurement error, but it is a form of measurement error whose properties and severity are understood. We all recognize that guessing may occur on any examination, and we all know how to discount for the possible impact of guessing. For example, when we administer to students a 100-item multiple-choice exam on which each question has four choice options, the expected value is 25, not zero.[4] Guessing is the devil we know. In contrast, there exists no direct means to ascertain the magnitude of contamination resulting from survey protocols that encourage DKs, nor is it possible to implement mathematical corrections to compensate for the loss of validity.

Finally, we doubt that truly blind guessing is highly prevalent on political knowledge questions.

---

with three choice options; thus, blind guessing would result in an average of 2.16 correct answers.

[4] Although the expected value in this example is 25, scores, at least in theory, can range from zero to 100. In reality, however, blind guessers will cluster close to 25. On a perfect 100-item multiple-choice test (i.e., no items contain contextual cues, and thus blind guessers have an equal probability of selecting each choice option) on which no test taker has any knowledge, the average number of questions answered correctly will be 25. It is possible, but extremely unlikely, for a test taker to receive a score of zero, and even more unlikely for a test taker to receive a score of 100. Specifically, the probability of answering every item incorrectly is 0.75^{100}, and the probability of answering every item correctly is 0.25^{100}. These are slim odds. We constructed a simulation with 100 items, and then ran the simulation 100,000 times. The minimum attained score was 8, the maximum score was 46, and the standard deviation was 4.35. The formula for variance in this case is np(1 - p), or 100(.25)(.75), producing an estimated standard deviation of 4.33, with a mean of 25. With three choice options, the mean score would be 33, and the

Informed guessing occurs more frequently on educational tests than does blind guessing (Mehrens and Lehmann 1984), and we suspect that the same is true for survey questions about politics. In other words, we expect that a large portion of respondents who are not fully informed on a subject are at least partially informed. Delli Carpini and Keeter report clear evidence on this point. They find that "citizens often approach the correct answer even if they do not demonstrate full command of it" (1996, 96).

Variance on knowledge scales will decrease when knowledge is measured using the procedures advocated here because systematic error variance will be minimized.[5] Consider, for example, a case in which respondents answer a five-item knowledge battery on which each question uses a three-category multiple-choice format, no respondent knows the answer to any item, 80 percent of respondents guess on each item, and 20 percent answer DK. In this scenario, the mean for 80 percent of respondents is 1.67, whereas the mean for the other 20 percent of respondents is zero. Scores for these latter respondents are pushed away from the mean as a consequence of the refusal guess, resulting in inflated variance. This example demonstrates the minimum exaggeration of variance that occurs with a guessing response set. On actual knowledge tests—on which at least some respondents possess full or partial information—the effect will be greater, because the mean will be higher and thus the refusal to guess will push respondents' scores an even greater distance from the mean.

We advocate designing measures of political knowledge in a manner consistent with established norms in education testing. We reiterate that our motivation in offering this recommendation is to improve the validity with which knowledge is measured, not to alter the substantive results. Nonetheless, likely changes in substantive findings should not be ignored. Error stemming from efforts to encourage DKs is systematic, not random. No respondent's political knowledge is overestimated because DKs are encouraged, but knowledge for many respondents may be underestimated. Consequently, the consensus in political science regarding the impoverished information base of the American electorate may

---

estimated standard deviation is 4.71.

[5] Variance will decrease for knowledge scales, but not necessarily for individual items. On highly difficult items,

overstate matters. How political scientists commonly ask knowledge questions may inadvertently cast respondents in an unduly negative light. We explore this matter below by using data from two split-ballot studies to contrast knowledge levels as typically measured in survey research and knowledge levels as measured using the protocol outlined here.

Data and Method

Data are drawn from two sources. The first is a survey we designed and conducted in the Tallahassee, Florida metropolitan area. As a complement to these data, we also placed a brief replication of the Tallahassee survey on the 1998 NES Pilot Study. Interviewing for the Tallahassee survey began in mid-November, 1998 and ended in late January, 1999. This survey was completed by 404 residents of Leon County, with interviews conducted by telephone. The mean age of respondents is 41 years; 54.0 percent of respondents are female; 71.6 percent of respondents are white and 21.3 percent are black.[6] The survey includes 98 items, and interviews took an average of 23 minutes to complete.[7] Interviewing for the 1998 NES Pilot was conducted from September 8 to November 3, with approximately 400 respondents drawn from each of California, Georgia, and Illinois. The mean age of respondents is 43 years, and 56.8 percent of respondents are female.[8]

---

discouraging DKs will increase variance.

[6] The estimated population of Leon County (1996) is 220,000; 72.3 percent of residents are white and 24.6 percent black.

[7] The ratio of completed interviews to refusals is exactly two to one. We calculate the completion rate at 49.6 percent. This figure is derived by dividing completions by the sum of completions, refusals, and unresolved cases. Unresolved cases are ones in which the telephone number was known to be that of a household, but where we were unable to obtain a completion or a refusal in twelve calls to the number. We also have a large number of undetermined cases. These are cases in which we are not certain that the telephone number is that of a household, and no contact was made with a person at the number in twelve calls. The number of undetermined cases is unusually high because we were not able to obtain a pre-cleaned sample (i.e., telephone numbers known to be working numbers at households) for the Tallahassee metropolitan area. We could have obtained a pre-cleaned sample for a different metropolitan area, or for the state of Florida as a whole. However, we simply did not have the resources available for the purchase of such a sample or for long-distance telephone calls. When undetermined cases are added to the denominator, the completion rate drops to 31.5 percent. Although we would have preferred a higher completion rate, our primary concern was with obtaining a sample sufficiently large to enable meaningful comparison of the two versions of our knowledge items. All comparisons discussed below involve assessment of one knowledge battery relative to another (i.e., our procedure is comparable to a laboratory experiment). Our sample is more than adequate for these purposes.

[8] The NES Pilot does not include a standard measure of race; instead, respondents are asked to identify ethnic

Analysis will focus on 15 knowledge items that were asked as part of the Tallahassee survey, and four items on the NES Pilot. We used a split-ballot design in each case, with respondents assigned randomly to one of two versions of the knowledge batteries. The first version was designed in a manner consistent with current convention in political science in that no effort was made to discourage DKs. Specifically, interviewers were instructed not to prompt for substantive answers when respondents indicated uncertainty, and Delli Carpini and Keeter's (1996, 305) recommended introduction was read to respondents: "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." The second version, in contrast, was designed in a manner consistent with current convention in educational testing. Interviewers were instructed to prompt respondents for substantive answers, and the introductory statement explicitly discouraged DKs: "Many people don't know the answers to these questions, but even if you're not sure I'd like you to tell me your best guess." Subtle changes in the wording of some of the questions on surveys further encouraged or discouraged DKs. On item #1 in Tallahassee, for example, respondents on version one were asked "Do you know who has the final responsibility to decide if a law is Constitutional or not? Is it the President, Congress or the Supreme Court?" In contrast, respondents on version two were asked "Who has the final responsibility to decide if a law is Constitutional or not? Is it the President, Congress or the Supreme Court?"

Knowledge is initially operationalized as a simple count of correct answers. Using data from version two of each survey, we also computed a second score with added corrections. First, respondents who initially answered “don’t know” on version two of the NES Pilot were prompted in follow-up items to offer substantive responses. Where substantive answers were given on these follow-up items, we use these answers rather than the initial DKs. Second, we eliminated any remaining DKs for version two of the NES and all DKs obtained on the multiple-choice items on version two in Tallahassee by randomly assigning them to the substantive response categories. In essence, this procedure means that we entered blind guesses for those respondents who refused to do so on their own. The benefit of this procedure is

---

groups with which they identify.

that respondents now are not punished for declining to answer; consequently, variance in "knowledge" will no longer be affected by an irrelevant property, the differential propensity to guess (Kline 1986).

Our third correction accounts for possible partially-informed responses on the Tallahassee survey. Past NES surveys have asked respondents to identify political figures such as the Speaker of the House and the Chief Justice of the Supreme Court. Four items of this type were asked on the Tallahassee survey. Research in educational testing confirms the intuition that such items are more difficult to grade objectively than are questions that use a multiple-choice format (Aiken 1986; Ebel and Frisbie 1986; Mehrens and Lehmann 1984; Stanley and Hopkins 1972). The problem concerns how to score answers that are partially correct, such as "Al Gore works with Bill Clinton," "and "William Rehnquist is on the Supreme Court." Typically, incomplete and partially-correct responses are counted as incorrect, and thus the survey respondent receives zero credit. We find this to be problematic. Just as a teacher might award students partial credit in some instances, we sought to do the same with data from the Tallahassee survey. Nearly five percent of answers on the identification items were deemed by interviewers to be partially correct. To determine how much credit to award for partially-correct responses, we created a count of correct answers on the eleven multiple-choice items, and then regressed this variable on counts of correct and partially-correct responses on the identification questions. This procedure yielded results indicating that each partially-correct response should be awarded 0.725 points.[9]

Results

Data from the two split-ballot surveys permit comparison of knowledge levels as conventionally measured in political science and as measured in accordance with established norms in the field of

[9] The regression model allows us to quantify approximately how much information is reflected by a partially-correct response relative to a fully-correct response. The dependent variable in this model is the number of multiple-choice items answered correctly, out of eleven. The first independent variable is the number of identification items answered correctly, and the second independent variable is the number of identification items for which the respondent received partial credit. The model produced a coefficient of 1.60 for b1, and 1.16 for b2, with both coefficients significant at $p < .001$. The ratio of the latter coefficient to the former is 0.725, meaning that a partially-correct response is weaker than a fully-correct response as a predictor of "knowledge" (as measured using the other available items). Using a similar procedure, we determined that a response volunteered by 16 people on item #7—that it was not possible to answer the question based on the information provided—should be awarded full credit.

educational testing. The first step is comparable to the manipulation check in a laboratory experiment. Given that the procedures used in version two of the split-ballot were designed to drive down the DK rate, we must determine whether this objective was achieved. The Tallahassee survey includes 11 multiple-choice items and four identification items. On version one, the overall DK rate was 20.0 percent, and the DK rate on the multiple-choice questions was 14.6 percent. On version two, the overall DK rate was 10.5 percent, and the rate was only 3.9 percent on the multiple-choice items. Hence, our procedures did not eliminate all DKs, particularly on the four identification items, but the DK rate was reduced substantially relative to version one.

Results are similar on the NES Pilot. On version one, the DK rate on the four multiple-choice items was 20.5 percent. This was reduced to 5.7 percent on version two, and to 3.4 percent with efforts to convert DKs into substantive answers. Together, data from the two surveys reveal that relatively minor changes in the phrasing of the introductory statement and in question wording were sufficient to drive out nearly three-quarters of would-be DKs. Given the threat to validity associated with encouraging DKs, the new protocol tested here represents a cost-free means to improve the quality of knowledge scales.

In numbers, the reduction in the overall DK rates suggests that efforts to discourage DKs resulted in the conversion of approximately 285 would-be DKs into substantive answers in Tallahassee, and 365 on the NES Pilot.[10] Conventional knowledge scales code DKs and incorrect answers identically, with each receiving a score of zero to indicate the absence of knowledge. Hence, DKs and incorrect answers are viewed as interchangeable; they are two ways of denoting the same thing, namely that the respondent does not know the correct answer. It follows that if encouraging DKs is harmless, then all of the erstwhile DKs in version two of our surveys should turn up as incorrect answers. This did not occur. In Tallahassee, we estimate that two of every three converted DKs was answered correctly in version two, resulting in nearly one additional correct answer per respondent (the proportion of items answered

[10] We calculated these estimates by projecting DKs in version two of each survey using the DK rate observed on version one, and comparing these projections with observed DK levels on version two.

correctly in version two was 66.3, versus 60.0 in version one; the proportion incorrect was 23.2, versus 20.0). On the NES, our estimate is that over 53 percent of converted DKs were answered correctly.

These results warrant closer inspection. Question wording and rates of correct response for the 15 knowledge items asked in Tallahassee are reported in table 2. The first 11 items use a multiple-choice format. Items 3, 4, and 5 each have two choice options, whereas the remaining multiple-choice questions include three choice options. Questions 12-15 are open-ended identification items. Questions 1-6 and 12-15 are drawn from previous surveys. Questions 7-11 are new items designed specifically for the Tallahassee project. Critics of conventional knowledge items argue that survey researchers have focused undue attention on the ability of respondents to recall sometimes obscure facts about politics (e.g., Graber 1996; Neuman, Just and Crigler 1992). We drafted items 7-11 in response to this criticism. The content in each case centers more on conceptual understanding than on simple recall.[11]

In table 2, column one depicts results for version one of the split-ballot, in which DKs were encouraged. Column two depicts the raw data for version two, in which DKs were discouraged. Column three adds our corrections to data from version two: DKs on the multiple-choice items are assigned randomly to the substantive choice categories, and credit is awarded for partially-correct responses. Comparing column one and column two, we see that respondents in version two of the split-ballot outperformed their counterparts in version one on every item except for item 6. Thus, the general increase in knowledge scores described above is not the consequence of just a handful of items; across the board, different results are obtained when DKs are discouraged rather than encouraged.

Open-ended items are not susceptible to blind guessing. For instance, the odds are rather slim that a person who has absolutely no idea what office Trent Lott holds would come up with Senate

[11] This is a first take at measuring conceptual knowledge, and thus we make no claims regarding the virtues of these specific items. However, the use of a three-category multiple-choice format was deliberate. Writers in educational testing agree that a short-answer format restricts content to straightforward factual material (e.g., "What job or political office is currently held by Trent Lott?"), whereas multiple-choice items allow the examiner to tap more conceptual subject material (Aiken 1986; Ebel and Frisbie 1986; Mehrens and Lehmann 1984; Stanley and Hopkins 1972). Additionally, research in educational testing indicates that multiple-choice questions perform best when three choice options are included (e.g., Haladyna and Downing 1993; Landrum, Cashin and Theis 1993; Tversky 1964).

Majority Leader by sheer chance. With this in mind, results for items 12-15 warrant emphasis. Although only a relatively few DKs were discouraged by the procedures used in version two (the DK rate for items 12-15 is 34.6 percent in version one and 28.5 percent in version two), virtually all of the would-be DKs were converted into correct answers. It appears that open-ended questions asked in a format with a strong DK option systematically understate knowledge levels because some people who *do* know the answers respond that they do not.

Subjective scoring constitutes a second problem with the identification items. Interviewers determined that nearly half of the "incorrect" answers on the open-ended items on version two deserved credit as being partially correct. Partially-correct responses were especially common on the questions concerning Trent Lott and William Rehnquist. On conventional knowledge scales, respondents are awarded identical credit—none—regardless of whether they identify Trent Lott as "a leading Republican," "a Senator from Mississippi," "the vice president," or "an actor from that new police show." Results in column three demonstrate that awarding partial credit when appropriate yields substantially higher scores on these questions. Open-ended identification items pose a triple threat to accurate measurement of political knowledge: these questions inherently tend to focus on factual rather than conceptual subject matter; they generate high DK rates that mask substantive understanding; and they require subjective scoring that, at least as performed on most surveys in political science, ignores partial knowledge.

Results for the four NES items are reported in table 3. Data are highly similar to those from the Tallahassee survey in that efforts to discourage DKs yielded higher rates of correct response for all four NES items, and the corrections we add in column three produced further gains. These data also provide evidence regarding a benefit of the three-category multiple-choice format. Differences between column two and column three are trivial for the first two items, which use the three-category format, but gaps are larger for the third and fourth items, which use a two-category format. A similar pattern is evident in table 2 (items 1-2 and 6-11 have three choice options; items 3-5 have two choice options). Post-survey

corrections are minor with the three-category items because these questions elicited relatively few DKs, and because random assignment of DKs, like blind guessing, has only a one-in-three chance of generating a correct response. Put differently, the use of three choice categories rather than two is preferable both because fewer post hoc corrections are needed when three categories are used and because increasing the number of choice options decreases the negative effects of blind guessing.

We hypothesized that encouraging DKs understates levels of political knowledge, and overstates levels of variance. Results for the 19 individual items appear to be consistent with at least the first of these hypothesizes, but more systematic assessment requires that we draw on these data to construct multi-item scales. Table 4 reports results for five scales using the Tallahassee data: 1) all 15 items, 2) the five items recommended by Delli Carpini and Keeter, 3) the eleven multiple-choice items, 4) the four open-ended items, and 5) the five conceptual items. The sixth scale in table 4 is constructed using data from the four NES items. Results are consistent across the various scales. First, knowledge levels in version two exceed those in version one by statistically significant margins for all six scales, and these margins increase when corrections are added in column three. Second, the largest standard deviation for each of the six scales is obtained when DKs are encouraged.[12]

The measurement of political knowledge holds most true to the recommendations in educational testing when DKs are discouraged, observed DKs are assigned randomly to available choice options, and scoring on open-ended questions awards credit for partially-correct responses. Hence, comparison of the first and third columns in table 4 indicates the difference between following the conventional measurement strategy in political science and adhering to accepted testing practices. Means in column three exceed those in column one by from 10.0 percent (the Delli-Carpini and Keeter five-item scale) to

[12] To maximize clarity in presentation, we have analyzed all data from the NES Pilot collectively. However, because the Pilot includes independent samples drawn from three states (California, Illinois, Georgia), additional information can be obtained by analyzing data separately by state, thereby providing three independent replications. When we conducted the analysis by state, the key results for each state were identical to those reported in table 4: 1) statistically significant increases in knowledge levels comparable to those reported in column two and column three of table 4 were found in each state, and 2) the highest standard deviation in each state was obtained using data from version one of the split-ballot. The only difference in results by state is that mean scores in Georgia are lower than

16.7 percent (the NES scale). For the full 15-item scale in Tallahassee, the mean in column three is 15.3 percent larger than the mean in column one. Similar results are obtained upon comparison of levels of variance. The standard deviation for the open-ended items is only slightly smaller in column three than in column one (6.7 percent). Larger discrepancies emerge for all other scales. The standard deviations for the Delli-Carpini and Keeter scale differ in magnitude by 15.6 percent; the gaps are between 13 and 14 percent for both the full 15-item Tallahassee scale and the four-item NES measure.

Implications

Two common conclusions regarding the distribution of political knowledge—namely that the mean is low, and the variance high—are fueled by widespread use of suboptimal measurement procedures. Questions designed to measure political knowledge elicit high "don't know" rates, and many of these DKs mask substantive understanding. When survey procedures follow established tenets in the field of educational testing, results reveal that knowledge is not as low, and variance in knowledge is not as high, as previously thought. Specific findings indicate that encouraging DKs systematically skews the key statistics by a minimum of 10 to 15 percent.

A fair question concerns whether the changes brought by use of corrected measures are of sufficient magnitude for us to reconsider fundamental conclusions regarding the distribution of information among American citizens. One counter to our findings is that shifts of 10 to 15 percent are minor, and thus do not alter the basic fact that Americans know little about politics and government. A second counter is that data such as those we present are largely beside the point. Anecdotal evidence abounds regarding the political ignorance of the American public, and the data presented here do not refute the claim that knowledge levels are low.[13]

---

those in California and Illinois for all versions of the knowledge scales.

[13] An additional point that warrants mention is whether we should be concerned at all about levels of political knowledge. Most research on political knowledge is intended to identify the antecedents and consequences of variance in knowledge levels, not to describe how much Americans know about politics. Hence, perhaps attention to levels of knowledge is misguided. We disagree with this view. It is precisely because of the low mean and high variance in knowledge that political scientists have devoted so much attention to studying knowledge. Researchers have attempted to determine the effects of an uninformed citizenry, and the effects of wide disparities in knowledge

These are reasonable objections. We possess no guideline regarding what Americans should know about politics, nor about how much knowledge is enough. When we designed our eleven multiple-choice questions in accordance with recommendations from the field of educational testing, Tallahassee respondents, on average, answered 70 percent correctly. Viewed in absolute terms, this clearly is not a sterling performance. However, we believe that results should be assessed relative to our prior understanding about the distribution of political information. From this perspective, several points merit discussion.

First, independent of substantive conclusions, the implications of our findings for the measurement of political knowledge should not be ignored. Ideally, observed scores on knowledge scales should be a function of one systematic factor (knowledge) and one unsystematic factor (chance). The common practice in political science of encouraging DKs introduces a second systematic source of variance (propensity to guess), thereby limiting validity. Research in educational testing provides a means to increase the validity of knowledge measures, and we see no justification to reject this opportunity. Surveys should discourage, not encourage, DKs; when DKs are recorded on multiple-choice items, those answers should be assigned randomly to the available choice options; and, given the numerous problems involved with open-ended items in terms of validity, limited substantive content, and ambiguous coding, use of these items should be avoided.

Second, changes in means and standard deviations of 10 to 15 percent are not trivial. We concede that this ultimately is a matter of opinion, yet statistical shifts of the magnitude reported here certainly would draw attention were the subject matter to concern topics such as life expectancy, annual pay raises,

---

levels. If all Americans were highly informed about politics, knowledge would be much less interesting to political scientists. Consequently, our finding that both the depth of political ignorance and the variance in knowledge levels have been exaggerated by faulty measurement procedures speaks directly to research on the antecedents and consequences of variance in knowledge. In related research, we explore whether the impact of knowledge as an independent variable changes when knowledge is measured in accordance with recommendations from the field of educational testing. We find that it does. The short story is that the effect of knowledge weakens with valid measures in cases in which the dependent variable is prone to exaggeration by respondents (e.g., self-reported voter turnout and political participation), whereas the effect of knowledge often strengthens considerably with valid measures when knowledge is used as part of an interactive specification (e.g., whether the impact of ideology on political

job openings in academia, presidential approval, or a House member's margin of victory. Imagine walking into the classroom and announcing "it turns out that I have been making a systematic error in calculating your grades. On average, your actual score in the class is approximately one full letter grade higher than I reported previously." We're betting that your students would find this to be a meaningful change.

Third, relevant evidence should matter for the conclusions we reach. We appreciate that many political scientists who have commented on Americans' levels of political knowledge have been motivated by normative considerations. Nonetheless, the data should not be ignored. Political scientists concur that Americans possess a "paucity of information" about politics. Well, how do *we* know this? In large part, we know this because Americans have fared poorly on questions about politics asked on our surveys, but those questions inadvertently have stacked the deck against the American citizen. If evidence informs our conclusions, then our conclusions now must be restated in at least somewhat less extreme terms. In the past, we have encouraged people to say "I don't know," and then, when they did so, we have fretted over their ignorance. This state of affairs is partly of our own making. When questions are asked *and* answered, it turns out that Americans know more about politics than we thought.

References

Aiken, Lewis R. 1988. *Psychological Testing and Assessment*, 6th ed. Boston: Allyn and Bacon.

---

judgments varies as a function of the respondent's knowledge level).

Ben-Shakhar, G., and Y. Sinai. 1991. “Gender Difference in Multiple-Choice Tests: The Role of Differential Guessing Tendencies.” *Journal of Educational Measurement* 28:23-35.

Brown, Frederick G. 1970. *Principles of Educational and Psychological Testing*. Hinsdale, Ill.: The Dryden Press Inc.

Casey, M. Beth, Ronald L. Nuttall, and Elizabeth Pezaris. 1997. "Mediators of Gender Differences in Mathematics College Entrance Test Scores: A Comparison of Spatial Skills with Internalized Beliefs and Anxieties." D*evelopmental Psychology* 33:669-80.

Converse, Philip E. 1990. “Popular Representation and the Distribution of Information.” In *Information and Democratic Processes*, ed. John A. Ferejohn and James H. Kuklinski. Urbana: Universit of Illinois Press.

Cronbach, Lee J. 1950. “Further evidence on Response Sets and Test Design.” *Educational and Psychological Measurement* 10:3-31.

Cronbach, Lee J. 1946. “Response Sets and Test Validity.” *Educational and Psychological Measurement* 6:475-94.

Cronbach, Lee J. 1942. “Studies of Acquiescence as a Factor in the True-False Test.” *Journal of Educational Psychology* 33:401-15.

Cunningham, George K. 1986. *Educational and Psychological Measurement*. New York: Macmillan Publishing Company.

Delli Carpini, Michael X., and Scott Keeter. 1996. *What American Know About Politics and Why It Matters*. New Haven, Conn.: Yale University Press.

Delli Carpini, Michael X., and Scott Keeter. 1993. "Measuring Political Knowledge: Putting First Things First." *American Journal of Political Science* 37:1179-1206.

Ebel, Robert L., and David A. Frisbie. 1986. *Essentials of Educational Measurement*, 4th ed.

Ferejohn, John A. 1990. “Information and the Electoral Process.” In *Information and Democratic Processes*, ed. John A. Ferejohn and James H. Kuklinski. Urbana: Universit of Illinois Press.

Gritten, Frances, and Donald M. Johnson. 1941. “Individual Differences in Judging Multiple-Choice Questions.” *Journal of Educational Psychology* 32:423-30.

Haladyna, Thomas M., and Steven M. Downing. 1993. “How Many Options is Enough for a Multiple-Choice Test Item?” *Educational and Psychological Measurement* 53:999-1010.

Hirschfeld, Mary, Robert L. Moore, and Eleanor Brown. 1995. "Exploring the Gender Gap on the GRE Subject Test in Economics." *Journal of Economic Education* 26:3-15.

Kline, Paul. 1986. *A Handbook of Test Construction*. London: Methuen & Co.

Landrum, R. Eric, Jeffrey R. Cashin, and Kristina S. Theis. 1993. “More Evidence in Favor of Three-Option Multiple-Choice Tests.” *Educational and Psychological Measurement* 53:771-78.

Mehrens, William A., and Irvin J. Lehmann. 1984. *Measurement and Evaluation in Education and Psychology*, 3rd ed. New York: Holt, Rinehart and Winston.

Mondak, Jeffery J. 2000. “Reconsidering the Measurement of Political Knowledge.” *Political Analysis* 8:57-82.

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

Nunnally, Jum. 1972. *Educational Measurement and Evaluation*, 2nd ed. New York: McGraw-Hill Book Company.

Page, Benjamin I., and Robert Y. Shapiro. 1992. *The Rational Public: Fifty Years of Trends in Americans' Policy Preferences*. Chicago: University of Chicago Press.

Popkin, Samuel L. 1994. *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*, 2nd ed. Chicago: University of Chicago Press.

Sherriffs, Alex C., and Donald S. Boomer. 1954. "Who Is Penalized by the Penalty for Guessing?" *Journal of Educational Psychology* 45:81-90.

Slakter, Malcolm J. 1969. "Generality of Risk Taking on Objective Examinations." *Educational and Psychological Measurement* 29:115-28.

Slakter, Malcolm J. 1968a. "The Effect of Guessing Strategy on Objective Test Scores." *Journal of Educational Measurement* 5:217-22.

Slakter, Malcolm J. 1968b. "The Penalty for Not Guessing." *Journal of Educational Measurement* 5:141-44.

Sniderman, Paul M., Richard A. Brody, and Philip E. Tetlock. 1991. *Reasoning and Choice: Explorations in Political Psychology*. New York: Cambridge University Press.

Stanley, Jullian C., and Kenneth D. Hopkins. 1972. *Educational and Psychological Measurement and Evaluation*. Englewood Cliffs, N.J.: Prentice-Hall.

Swineford, F. 1941. "Analysis of a Personality Trait." *Journal of Educational Psychology* 32:438-44.

Thorndike, Robert L., and Elizabeth Hagen. 1969. *Measurement and Evaluation in Psychology and Education*, 3rd ed. New York: Wiley.

Tversky, Amos. 1964. "On the Optimal Number of Alternatives at a Choice Point." *Journal of Mathematical Psychology* 1:386-91.

Wiley, Llewellyn, and Otis C. Trimble. 1936. "The Ordinary Objective Test as a Possible Criterion of Certain Personality Traits." *School and Society* 43:446-48.

Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.

Table 1. Possible Relationships between Underlying Knowledge States and Responses to Multiple-choice Questions

	Answer Recorded on Survey		
Underlying Behavioral State	Correct Response	Incorrect Response	"Don't Know"
Fully Informed	✓ respondent knows answer and offers a substantive response	not possible	✓ respondent knows answer, but declines to offer a substantive response
Partially Informed	✓ respondent is less than fully certain of the correct answer, but chooses the correct option	✓ respondent is less than fully certain of the correct answer, and chooses an incorrect option	✓ respondent is less than fully certain of the correct answer, but declines to offer a substantive response
Misinformed	not possible	✓ respondent wrongly believes he or she knows the correct answer and offers a substantive response	✓ respondent wrongly believes he or she knows the correct answer but declines to offer a substantive response
Uninformed	✓ respondent possesses no information pertinent to the question, but chooses the correct option by chance	✓ respondent possesses no information pertinent to the question, and chooses an incorrect option by chance	✓ respondent possesses no information pertinent to the question, and declines to offer a substantive response

Table 2. Correct Responses on Tallahassee Knowledge Items
(percentage answering correctly)

Item	"Don't Know" Responses Encouraged	"Don't Know" Responses Discouraged: No Added Corrections	"Don't Know" Responses Discouraged: Corrections for DKs, partial information
1) Who has the final responsibility to decide if a law is Constitutional or not? Is it the president, Congress, or the Supreme Court?	59.8	70.0	70.0
2) Whose responsibility is it to nominate judges to the Federal Courts? The president, Congress, or the Supreme Court?	55.4	61.5	62.5
3) Which party had the most members in the House of Representatives in Washington before the election this month?	69.6	78.5	80.0
4) Which party had the most members in the U.S. Senate before the election this month?	59.8	75.0	77.5
5) Which one of the parties is more conservative than the other at the national level, the Democrats or the Republicans?	70.4	77.4	79.9
6) How much of a majority is required for the U.S. Senate and House to override a presidential veto—a bare majority (50% plus one), a two-thirds majority, or a three-fourths majority?	73.5	72.0	73.0
7) Suppose that a liberal senator and a conservative senator must vote on a bill that would close down a federal regulatory agency. Which senator is more likely to vote to close down the agency—the liberal senator, the conservative senator, or would you say that both senators are equally likely to vote to close the agency?	40.1	47.7	51.8
8) Which of the following does the U.S. government do to limit criticism of the government in newspapers? Does the government require newspapers to be licensed; does the government prohibit stories that are too critical of the government from being published; or would you say that the U.S. government has no regulations that limit political criticism in newspapers?	77.9	83.0	84.5
9) Which of the following is the main duty of Congress—to write legislation; to make sure that the president's policies are carried out properly; or to watch over the states' governments?	62.3	68.5	70.0

10) How does the political system in the United States try to make sure that political leaders are accountable to citizens—by having elections; by limiting the president to two terms in office; or by having separate state courts and federal courts?	64.5	70.5	71.5
11) In some countries, voters elect leaders who serve in a parliament. Is a parliamentary government authoritarian, autocratic, or democratic?	41.7	43.9	48.5
12) What job or political office is currently held by Al Gore?	87.3	93.0	93.4
13) What job or political office is currently held by Newt Gingrich?	72.9	79.0	81.8
14) What job or political office is currently held by Trent Lott?	31.4	35.5	40.7
15) What job or political office is currently held by William Rehnquist?	33.5	38.5	44.9
Maximum Number of Cases	204	200	200

Source: 1998-1999 Tallahassee Survey
Note: In column one, the knowledge battery began with the following preface: “Now we have 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.” In column two and column three, the battery began with this preface: “Here are a few questions about the government in Washington. Many people don’t know the answers to these questions, but even if you’re not sure I’d like you to tell me your best guess.” The same data are used in column two and column three. Two corrections are made in column three: 1) “don’t know” responses on the objective items (1-11) are randomly assigned to the substantive response categories to mimic blind guessing; and 2) partial credit is given on item 7 and items 12-15 if the respondent’s answer showed correct but incomplete understanding of the concept or person in question (e.g., identifying William Rehnquist as a member of the Supreme Court, but not as Chief Justice.

Table 3. Correct Responses on NES Pilot Knowledge Items
(percentage answering correctly)

Item	"Don't Know" Responses Encouraged	"Don't Know" Responses Discouraged: No Added Corrections	"Don't Know" Responses Discouraged: Corrections for DKs
1) Who has the final responsibility to decide if a law is Constitutional or not? Is it the president, Congress, or the Supreme Court?	54.3	61.5	61.8
2) And whose responsibility is it to nominate judges to the Federal Courts? The president, Congress, or the Supreme Court?	47.3	53.2	53.6
3) Which party has the most members in the House of Representatives in Washington?	66.3	72.1	76.7
4) Which party has the most members in the U.S. Senate?	57.2	64.6	70.8
Maximum Number of Cases	582	620	620

Source: 1998 NES Pilot
Note: In column one, the knowledge battery began with the following preface: "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." In column two and column three, the battery began with this preface: "Here are a few questions about the government in Washington. Many people don't know the answers to these questions, but even if you're not sure I'd like you to tell me your best guess." The same data are used in column two and column three. Two corrections are made in column three: 1) respondents who initially answered "don't know" were encouraged in a separate item to attempt to provide a substantive response, and correct answers on these follow-up items are included as correct responses in column three; and 2) respondents who answered "don't know" on both the initial item and the follow-up prompt were assigned randomly to the substantive response categories to mimic blind guessing.

Table 4. Alternate Versions of Knowledge Scales

Scale		"Don't Know" Responses Encouraged	"Don't Know" Responses Discouraged: No Added Corrections	"Don't Know" Responses Discouraged: Corrections for DKs, partial information
Full 15-item scale	Mean	8.944	9.923**	10.308***
	s.d.	3.862	3.504	3.353
	N	198	195	195
Delli-Carpini and Keeter five-item scale (items 1, 3, 5, 6, 12)	Mean	3.601	3.914**	3.960**
	s.d.	1.412	1.241	1.192
	N	203	198	198
Multiple-choice items (items 1-11)	Mean	6.725	7.469**	7.679***
	s.d.	2.902	2.612	2.465
	N	200	196	196
Open-ended identification items (items 12-15)	Mean	2.243	2.462*	2.611**
	s.d.	1.248	1.162	1.164
	N	202	199	199
Conceptual items (items 7-11)	Mean	2.866	3.147*	3.269*
	s.d.	1.424	1.311	1.291
	N	201	197	197
NES Pilot four-item scale	Mean	2.249	2.508***	2.624***
	s.d.	1.407	1.289	1.212
	N	582	620	620

Source: Table 2, Table 3
Note: asterisks in column two and column three indicate significant one-tailed t-tests in contrasts vs. column one;
*** $p < .001$; ** $p < .01$; * $p < .05$