Reconsidering the Measurement of Political Knowledge

Jeffery J. Mondak
Associate Professor
Department of Political Science
Florida State University
Tallahassee, FL 32306
850-644-7302
fax: 850-644-2367
jmondak@garnet.acns.fsu.edu

Acknowledgments

Mitch Sanders and John Zaller provided very helpful feedback on earlier drafts of this paper. Helpful comments also were received from Damarys Canache, Belinda Davis, Paul Goren, Carl McCurley, several anonymous reviewers, a Delta flight attendant, and Alex Korp and Elana Wakeman (from the crew of the movie "Rose's"), Some of the data utilized here were made available by the Inter-university Consortium for Political Science Research. The data were originally collected by the National Election Studies. Neither the collector nor the Consortium bears any responsibility for the analyses or interpretations presented here.

Abstract

Political knowledge has emerged as one of the central variables in political behavior research, with numerous scholars devoting considerable effort to explaining variance in citizens' levels of knowledge and to understanding the consequences of this variance for representation. Although such substantive matters continue to receive exhaustive study, questions of measurement also warrant attention. I demonstrate that conventional measures of political knowledge—constructed by summing a respondent's correct answers on a battery of factual items—are of uncertain validity. Rather than collapsing incorrect and don't know responses into a single absence-of-knowledge category, I introduce estimation procedures that allow these effects to vary. Grouped-data multinomial logistic regression results demonstrate that incorrect answers and don't knows perform dissimilarly, a finding that suggests deficiencies in the construct validity of conventional knowledge measures. The likely cause of the problem is traced to two sources: knowledge may not be discrete, meaning that a simple count of correct answers provides an imprecise measure; and, as demonstrated by a wealth of research conducted in the field of educational testing and psychology since the 1930s, measurement procedures used in political science potentially result in "knowledge" scales contaminated by systematic personality effects.

1. Introduction

Political knowledge stands as one of the central variables in contemporary research on mass political behavior. Unfortunately, it does so with unsure footing. It is common for researchers to ask survey respondents a battery of factual questions about politics, with political knowledge represented by a count of how many items a person answered correctly. I will show that this seemingly straightforward approach yields data of highly ambiguous meaning. Several researchers have reported works exploring the validity of knowledge measures, yet one crucial point has received insufficient attention: a simple count of correct answers provides a valid measure of knowledge only if it is appropriate to collapse the other two response categories—the incorrect answer and the "don't know"—into a single "absence of knowledge" grouping. I will demonstrate that this common practice is not appropriate.

The measurement of political knowledge seems simple at first glance. Correct answers represent knowledge, and thus more correct answers imply higher levels of knowledge; conversely, items not answered correctly indicate an absence of knowledge. This seemingly innocuous view implicitly introduces several demanding assumptions. First, when depicted as described here, knowledge scales enjoy near-perfect validity in the sense that correct answers essentially *mean* that a respondent holds the knowledge in question. Second, knowledge is a discrete property. For any question of fact, one either does or does not know the correct answer. There are no shades of gray. Third, knowledge scales measure only knowledge. No attribute other than knowledge varies systematically as a function of how many items respondents answer correctly. Curiously, these assumptions have aroused little concern in the past. Instead, researchers have focused on other important questions in the measurement of political knowledge, especially the issue of *which* items should be used (Delli Carpini and Keeter 1993, 1996), and whether knowledge tends to be general or domain-specific (Iyengar 1990; Zaller 1992).

The stakes are quite high because political knowledge has been advanced as our best measure of political sophistication (e.g., Delli Carpini and Keeter 1996; Luskin 1987; Neuman 1986) and political awareness (Zaller 1992)—pivotal concepts in research on mass politics. In these capacities, knowledge has seen frequent use as a dependent variable (e.g., Bennett 1988, 1989; Bennett and Bennett 1989; Cassel and Lo 1997; Delli Carpini and Keeter 1996; Lambert et al. 1988; Mondak

1995; Pettey 1988; Verba, Burns and Schlozman 1997; Weaver and Drew 1993), an independent variable (e.g., Delli Carpini and Keeter 1996; Neuman 1986; Price and Zaller 1993), and a moderating variable (e.g., Althaus 1998; Bartels 1996; Delli Carpini and Keeter 1996; Goren 1997; Zaller 1990, 1992). Given the centrality of knowledge scales in research on mass politics, we must not accept without question the validity of these measures. Thus, my purpose in this paper is to examine the critical assumptions both analytically and empirically. This paper will establish that a problem does, in fact, exist, and offer a preliminary diagnosis as to why this is the case.

Discussion will proceed in four parts. First, I will examine knowledge scales analytically. This treatment will explore possible stumbling points in the measurement of political knowledge, and describe possible effects of these obstacles on the empirical performance of knowledge scales. Second, knowledge scales will be used as dependent variables in a series of diagnostic tests to determine if, in fact, we have stumbled. In essence, I will replicate in abbreviated form past multivariate assessments of the antecedents of "knowledge." Then, using a more appropriate statistical technique, I will show how previous inquiries have supported conclusions that are of ambiguous meaning. Third, in similar fashion, knowledge scales will be used as independent variables, with voter turnout standing in, for illustrative purposes, as the dependent variable. Past research has shown that turnout increases as a function of knowledge. In reexamining this result, I demonstrate that the relationship between knowledge and turnout is not nearly so straightforward as suggested by previous analysts. Finally, preliminary evidence on the question of why knowledge scales perform imperfectly will be reported.

2. Potential Obstacles in the Measurement of Political Knowledge

To understand how problems may arise in the measurement of political knowledge, we must begin by examining our scales' basic features. Responses to political knowledge items fall into three categories: correct answers, incorrect answers, and don't knows (DKs). Knowledge scales typically define "knowledge" as the sum of a respondent's correct answers on a battery of items, meaning that incorrect answers and DKs are collapsed into one category. A correct answer receives a score of one, whereas both an incorrect answer and a DK receive scores of zero. A wrong answer thus is assumed to mean precisely the same thing as a DK—each indicates the absence of knowledge.

If wrong answers and DKs mean the same thing, then we should expect that they will behave the same way. Consider an example. Several studies have reported that education is a strong predictor

of political knowledge; as education goes up, so too does the number of items answered correctly on knowledge batteries. This is equivalent to saying that as education goes up, ignorance (or whatever we choose to label the lack of knowledge) goes down. Given two presumably identical measures of the lack of knowledge—the incorrect answer and the DK—it follows that as education goes up, the number of incorrect answers should go down, and the number of DKs should decline as well.

If knowledge is discrete, and if our scales measure only knowledge, then we would be right to collapse incorrect answers and DKs into one category. However, two possibilities call this practice into question. First, knowledge may not be a discrete property. This poses a problem if the wrong answer indicates a different state than the DK. Delli Carpini and Keeter (1996) distinguish between respondents who are *misinformed, uninformed,* and what I will label as *partially informed.* If these states differ, then the meaning of *informed* must be contextualized. For instance, transforming an individual from misinformed to informed may have different behavioral consequences than transforming this person from uninformed to informed.[i] A second potential problem with knowledge scales is that they may measure something else in addition to knowledge. That is, the scales may not tap only a single construct. Drawing on a wealth of research in the field of educational testing and psychology, I will discuss the possibility that answers on knowledge batteries reflect various personality traits of our respondents.

2.1. The Problem of Partial Information

The possibility that wrong answers indicate a different state than DKs constitutes the first threat to knowledge scales. The conventional method of measuring political knowledge implicitly assumes that knowledge is discrete; a person either does or does not hold the knowledge in question. We count the number of correct answers a respondent gave just as we would count the number of televisions a family has in their home, or the number of speeches a candidate delivered. Based on past research alone, this strategy seems questionable. Consider again Delli Carpini and Keeter's (1996) suggestion that it means something different to be misinformed as opposed to uninformed. Common scaling practices ignore this possibility: incorrect answers (indicating misinformation) and DKs (indicating a lack of information) are scaled identically. Intuitively, though, the process by which one person comes to be misinformed differs from the process by which another person remains uninformed. To be misinformed implies that exposure to information occurred and that the

processing and storage of that information were somehow flawed. To be uninformed, in contrast, implies that no information was received and stored. If misinformed and uninformed represent different behavioral states, then the antecedents of those states may differ. Likewise, the consequences of these different states may vary. To treat these states as identical brings risk that meaning is lost.

A further complication discussed by Delli Carpini and Keeter (1996) stems from the fact that some respondents exhibit partial understanding of the topic in question even though their knowledge is insufficient for them to receive credit for a correct answer. This scenario resembles the situation in educational testing in which a test-taker on a multiple-choice exam is not sure of the right answer, yet can rule out one or more of the incorrect choices (Mehrens and Lehmann 1984). From this perspective, we can distinguish *four* behavioral states: fully informed ("it is the Supreme Court's responsibility to determine whether a law is Constitutional"), partially informed ("it's either the president or the Supreme Court; I know it's not Congress"); misinformed ("that would be the president"), and uninformed ("someone does that? I had no idea!"). Knowledge scales currently ignore these shades of gray.

If we assume that knowledge items measure only knowledge, then this would imply that the uninformed select DK, the misinformed answer incorrectly, and the fully informed answer correctly. But what of the partially informed? Some partially-informed respondents will answer correctly and others will answer incorrectly, with the likelihood of a correct answer increasing as a function of the level of partial information. This means that respondents can give correct answers without being fully informed, and they can give incorrect answers without being misinformed. In this scenario, the incorrect answer represents a higher state of knowledge than does the DK; the former indicates either misinformation or partial information, whereas the latter indicates the absence of information.

Use of knowledge scales to represent political sophistication and political awareness reinforces this point. Discussions of these concepts emphasize their rich, multifaceted character. Political sophistication, for instance, implies a mix of interest and attentiveness toward politics, understanding of relevant issues and events, and cognitive ability (e.g., Neuman 1986). If political knowledge functions as our best indicator of a multidimensional construct, then it follows that at least some possibility exists that knowledge is measured imperfectly with a count of correct answers.

Were we to use facial expressions to measure mood, we would not simply count the number of times a person smiles, because collapsing all non-smiling expressions into one category would sacrifice too much meaning. Grouping incorrect answers with DKs may bring similar consequences.

Definitions of political sophistication and political awareness highlight the difference between attentiveness to politics and understanding of the information one has received. From this perspective, it may be that misinformed citizens rate high on attentiveness and low on understanding. Similarly, citizens who are partially informed may outscore the uninformed on both attentiveness and understanding. This means that treating the misinformed and the partially informed as identical to the uninformed may limit our capacity to differentiate among the key dimensions of the construct we seek to study. Converse (1964) recognized this point, and thus he devised an ordinal scale that awarded incorrect answers higher status than DKs. Luskin (1987) rejects this logic, arguing that the difference between DKs and incorrect answers stems from personality traits of respondents rather than from variance in sophistication. As we will see below, Luskin's point rings true when considered from the context of research on educational testing and psychology. However, we also will encounter empirical results consistent with Converse's view. For now I leave this as an open question.

The analysis thus far suggests that respondents may offer incorrect answers when they are either misinformed or partially informed, and that both of these states possibly are best placed between fully informed (the correct answer) and uninformed (the DK). In this scenario, collapsing incorrect answers and DKs into one category brings imprecision because respondents with medium levels of attentiveness or sophistication are clumped together with those on the very bottom. Fortunately, the solution in this case is straightforward: if it is improper to merge incorrect answers and DKs, then precision would be restored simply by disaggregating those response categories. Before we move to adopt this correction, however, we must note that a second, and more severe, problem also may exist.

2.2. A Single Construct?

The second, and more disturbing, concern is that knowledge scales may measure one or more properties other than knowledge. If the number of correct answers given on a knowledge battery taps something in addition to knowledge, what might that something be? One quite plausible answer is that responses on our knowledge scales reflect respondents' personality traits, particularly self-confidence,

competitiveness, and propensity to take risks.

To understand how knowledge scales can measure personality traits, it is useful to think through the logic of the survey respondent who is asked to answer a knowledge battery. Again, a respondent can offer one of three answers: a correct response, an incorrect response, or a don't know. We have no way to stop a person who is only partially informed on a subject, or even completely uninformed, from offering a substantive reply. That is, we cannot force the respondent to say "I don't know." When partially-informed and uninformed respondents give substantive answers, some of those answers will turn out to be correct. Generally, more correct answers will be recorded for respondents who are partially informed ("Al Gore? I think I heard he's the vice-president, or something like that"), and for items that use a multiple-choice format ("Which party controlled the House before the election? Well, it's either the Democrats or the Republicans"). As noted above, the number of correct answers need not be perfectly correlated with levels of full information, because respondents can answer items correctly even when they are not fully informed. Indeed, respondents can answer correctly and be fully *uninformed* if the correct answer is the consequence of a lucky blind guess.[ii]

If all respondents give substantive answers whenever they are at least partially informed, or whenever the format of the question means that there are only two or three possible answers, then we would have a simple case of across-the-board grade inflation. Trouble would arise in the measurement of knowledge, however, if some partially-informed and uninformed respondents give substantive answers, and get some of these items right, while other partially-informed and uninformed respondents say that they do not know the answers to these same questions. In this instance, two people with identical levels of actual political knowledge could produce quite different observed scores.

Zaller (1992, 335; see also Fiske, Lau and Smith 1990) writes that "alone among the possible measures of awareness, tests of political knowledge are relatively immune to a social desirability response set; that is, individuals cannot overstate their levels of information holding because they perceive that it is socially desirable to appear politically aware." Zaller is correct that self-reports of media exposure, attentiveness, and so on are susceptible to exaggeration by respondents, whereas the objective measure of political knowledge is not. I believe, though, that the issue of social desirability

bias also must be approached from the opposite direction: might some survey respondents *understate* their levels of information holding by saying "I don't know" when they actually possess partial, or even full, information? Fiske, Lau and Smith (1990) reject this possibility, arguing that it is unlikely that respondents would attempt to appear uninformed. However, several motivations other than deliberate subterfuge could prompt choice processes that understate knowledge.

What might compel survey respondents to choose DK when they are at least partially informed? First, in some cases, DKs may be encouraged by wording on our surveys. For instance, Delli Carpini and Keeter (1993, 1996) recommend a preface to knowledge batteries that they indicate is designed to deter guessing.[iii] While discouraging blind guessing, it is possible that such statements also deter some informed guessing. This implies that respondents who are motivated to follow the directions we give on our surveys—and who interpret those directions as meaning that substantive answers should be offered only when the respondents are highly certain that they are correct—may unwittingly understate their levels of knowledge. Again, this means that respondents with equal levels of actual knowledge may receive quite different scores when we tabulate the number of items they answered correctly.

Second, the DK could be a means for disinterested respondents to complete the survey more quickly. On educational tests, test-takers share a common incentive to try their best: their grades depend on it. As a result, absent guessing penalties, only a few foolish students consciously leave any items blank. No such incentive to do one's best exists for the political knowledge measure, whereas thinking about a question prolongs surveys that many respondents surely see as tedious. Consistent with this view, and in contrast with educational tests, it is common for respondents on knowledge batteries to choose DK even when there are only two or three response options (e.g., Delli Carpini and Keeter 1996, 94).[iv]

Third, and most importantly, a social desirability effect may compel some respondents to understate their levels of political knowledge. Let us suppose that no partially-informed respondents are deterred by a survey that discourages guessing, and that all respondents try their best to answer our knowledge questions. Trouble still might arise if respondents think differently about what is socially desirable. If some respondents view it as desirable *to appear highly informed*, these individuals will defer from choosing DK—especially if they are highly self-confident—because every

DK is an item that cannot be answered correctly. In contrast, if other respondents view it as desirable *not to appear misinformed*, these individuals will be likely to choose DK—especially if they are lacking in self-confidence—because every DK is an item that cannot be answered incorrectly.

The logic outlined here reflects more than speculation. There exists a vast literature in the field of educational testing and psychology concerning the factors other than achievement that may affect the number of items test-takers answer correctly on objective examinations. Decades ago, it was the norm to encourage students to leave items blank if they were not certain of the right answer, and then to subtract points for incorrect answers. But this norm was challenged as early as the 1930s, and a near-consensus developed by the 1950s that efforts to discourage guessing severely confounded objective tests due to their tendency to result in exams that measure test-takers' personality traits (Brown 1970; Cronbach 1942, 1946; Cunningham 1986; Ebel and Frisbie 1986; Mehrens and Lehmann 1984; Nunnally 1972).

Self-confidence is one trait that may be measured when a DK option is available. Confident test-takers tend to offer substantive answers rather than admit that they do not know the correct response (Casey, Nuttall and Pezaris 1997; Gritten and Johnson 1941; Hirschfeld, Moore and Brown 1995; Sheriffs and Boomer 1954; Stanley and Hopkins 1972). Because some items will be answered correctly even though full knowledge is lacking, test scores will be artificially inflated. Conversely, students lacking self-confidence often leave items blank even when they know, or are at least reasonably certain, of the correct answers. Hence, knowledge for these students is understated by a count of correct answers (Sheriffs and Boomer 1954; Stanley and Hopkins 1972). This latter phenomenon in part reflects a social desirability bias in that individuals who are unduly concerned with the impression they make on others are most likely to seek shelter in the "don't know" response (Aiken 1988; Sherriffs and Boomer 1954).

Other traits that may be measured by objective tests that discourage guessing include risk-taking (Cronbach 1946; Sherriffs and Boomer 1954; Slakter 1969), and competitiveness (Hirschfeld, Moore and Brown 1995). The link between performance on objective tests and personality is so well established that factual tests have long been recognized as a *means* to measure personality traits (e.g., Wiley and Trimble 1936). Indeed, Slakter (1969) argues that these tests measure certain personality traits more reliably than they measure achievement.

Once research demonstrated the flaws inherent in discouraging guessing, a new norm developed in the field of educational testing and psychology. It is now accepted (Cronbach 1946, 1950; Kline 1986; Mehrens and Lehmann 1984; Nunnally 1972) that examiners should encourage test-takers to answer all items, even if they are not certain of some (i.e., test-takers should "guess"). This protocol best ensures that all test-takers capitalize on their partial knowledge (Slakter 1968a, 1968b, 1969). Because a few test-takers still leave items blank despite efforts to discourage nonresponse, Kline (1986) recommends that blanks be randomly assigned to the available response categories, an approach that, by mimicking blind guessing, eliminates DKs prior to the calculation of students' scores. Cronbach (1946) explains that encouraging test-takers to answer all items introduces random error (the consequence of blind guessing), and thus decreases reliability, but that this increase in *unsystematic* variance is well worth the cost because encouraging guessing removes *systematic* personality effects, thereby increasing validity.[v]

Three choice options (correct, incorrect, DK), and four levels of information (fully informed, partially informed, misinformed, uninformed) produce a 12-cell grid. Two cells are empty: the fully informed will not answer incorrectly, and the misinformed will not answer correctly. All other cells are viable: 1) correct answers can represent full information, partial information (the correct informed guess), and the lack of information (the correct blind guess); 2) incorrect answers can represent misinformation, partial information (the incorrect informed guess), and the lack of information (the incorrect blind guess); 3) DKs indicate that the respondent elected not to offer a substantive reply, which can occur regardless of level of information. When DK is eliminated as an option, the erstwhile DK responses are allocated to the remaining two choice categories: fully informed test-takers answer correctly, the misinformed answer incorrectly, and the partially-informed and the uninformed get some right and some wrong, with success determined by the level of partial information and by chance.

In the classroom, the behavior of students reinforces the practice of excluding DKs because students recognize that it is in their interest to answer every item. Unfortunately, the opposite practice dominates survey research on political knowledge. Respondents are encouraged to choose DK, and the absence of a stake in the outcome of a survey further motivates respondents to select this option. As a result, DKs tend to outnumber wrong answers, a pattern that is shown below with

data from recent NES surveys. If this preponderance of DKs occurs because scales tap both knowledge and personality, then past research using these scales possibly has misattributed identified effects.[vi]

Social scientists have addressed the possibility that DKs tap psychological states unrelated to the phenomena of interest, but this research has focused on attitude measures (e.g., Francis and Busch 1975; Rapoport 1979, 1982, 1985). The thinking that motivated these studies must now be used to guide reconsideration of the measurement of political knowledge. Interestingly, Rapoport (1979) advocates that DKs on attitude surveys be distributed among the substantive response categories, a correction identical to what Kline (1986) recommends for factual tests. Hence, where personality factors potentially confound our measures, one possible solution involves elimination of the offending choice category.

3. A Diagnostic Overview

The warning signs are of sufficient strength for us to doubt the precision of current knowledge scales, and NES data examined below provide additional grounds for concern. Analysis will center on the relationship between incorrect responses and DKs. Current scales will be called into question if data show that it is inappropriate to collapse these categories. To determine whether something is amiss, it will be necessary to assess the basic pattern of results. As an example, consider again the case in which education is used as a predictor of knowledge. Upon disaggregating DK and incorrect answers, we should find that education produces comparable effects on both indicators of the absence of knowledge. This pattern is shown in panel A of Figure 1, where the lines for incorrect answers and DKs run parallel to one another, meaning that the two response categories perform similarly.

Panel B and panel C depict examples in which DKs and incorrect answers behave dissimilarly. In both, the line for correct answers is identical to the one in panel A, meaning that the reported effect of education on "knowledge" would be the same in all cases were we to use current scaling practices. Panel B portrays a scenario in which education induces no variance in the number of incorrect responses; the increase in "knowledge" across education comes entirely at the expense of DKs. Panel C suggests an even more peculiar effect in that correct and incorrect answers both increase as a function of education, while DKs drop sharply.

If knowledge is not discrete, or if our scales are confounded by personality, disaggregation of

incorrect responses and DKs will produce patterns such as those in panels B and C. First, if incorrect answers reflect greater knowledge than DKs, then data for the three response categories should exhibit ordinality. Limiting analysis to a sum of correct answers would bring imprecision because two disparate slopes are merged—the weak effect of education on the propensity to offer correct rather than incorrect answers, and the strong effect of education on the propensity to offer correct answers rather than DKs.

Second, if knowledge scales are confounded by personality, then two distinct factors potentially distinguish the correct answer from the DK. Respondents may choose DK either because they truly do not know the answer or because they lack the confidence to offer a response. If predictors such as education are correlated with both knowledge and relevant personality traits, then patterns such as those depicted in panel B and panel C would emerge. The slope for incorrect answers would be between those for correct answers and DKs, but not because knowledge is ordinal. In this case, respondents who answer incorrectly resemble those who say DK in that knowledge is lacking, but they resemble those who give correct answers in that they possess sufficient self-confidence to offer substantive replies.

Similar discussion is warranted when knowledge is used as an independent variable. We know, for example, that political participation increases as a function of the number of knowledge items a person answers correctly. Suppose that we focus instead on the mirror opposite relationship, the negative effect of ignorance on participation. If participation declines as knowledge decreases, then we should observe a drop in participation as a function of increases in both incorrect answers and DKs. Once again, the critical diagnostic test is whether effects associated with incorrect answers closely resemble those associated with DKs.

4. Reexamining the Performance of Knowledge Scales

Data are drawn from the 1992, 1994, and 1996 NES surveys. I use these data because of their prominent role in research on American political behavior, and because these knowledge scales are similar to those used in most previous inquiries. The 1992 and 1994 scales include data from the nine available items, whereas seven items are used in 1996. Four short-answer identification items (the vice-president, the Chief Justice of the Supreme Court, the Russian president, and the Speaker of the House) are used in all years,[vii] as are items regarding party control of the House and the Senate, and

recognition that the Republican party is more conservative than the Democratic party. In 1992 and 1994, the two additional items concern whose responsibility it is to nominate federal judges, and whose responsibility it is to determine whether a law is Constitutional; these questions were not asked in 1996.

Four of the nine items used here (identification of the vice-president, party control of the House, more conservative party, judicial review) are among those recommended by Delli Carpini and Keeter as part of their five-item scale (1993, 1996).[viii] All nine of the items examined here performed reasonably well on Delli Carpini and Keeter's diagnostic tests. Hence, the current scales constitute fair representations of contemporary measures of factual political knowledge.

Descriptive data for the knowledge scales are shown in Table 1. The data reveal that the average respondent answered over half of the items correctly, and that DKs outnumber incorrect answers. The 1996 battery was a bit "easier" than the prior two, a finding that is of little surprise given developments in American politics between 1994 and 1996.[ix] Note that I have elected to treat the knowledge scales as proportions data; that is, rather than listing a simple count of how many responses appear in each of the choice categories, Table 1 reports these data as proportions. A cosmetic advantage accrues from treating the data in this manner, as comparison of scales with different numbers of items is made easier. A more tangible advantage is demonstrated below when the knowledge data are used as dependent variables.

Table 1. Correct, Incorrect, and “Don’t Know” Responses on Knowledge Scales: Descriptive Statistics

	1992	1994	1996
Number of items in scale	9	9	7
Mean proportion answered correctly (standard deviation)	0.514 (0.287)	0.540 (0.291)	0.643 (0.269)
Mean proportion answered incorrectly (standard deviation)	0.144 (0.133)	0.164 (0.135)	0.136 (0.143)
Mean proportion answered “don’t know” (standard deviation)	0.343 (0.281)	0.296 (0.262)	0.221 (0.245)
Number of Cases	2,244	1,771	1,480

Source: 1992, 1994, and 1996 National Election Studies.

Items used in the 1992 and 1994 knowledge scales are identification of Quayle, Rehnquist, Yeltsin, and Foley, identification of which party controlled the House before the election, identification of which party controlled the Senate before the election, identification of the more conservative political party, identification of responsibility for determination of whether a law is Constitutional, and identification of responsibility for nomination of federal judges. Items used in the 1996 knowledge scale are identification of Gore, Rehnquist, Yeltsin, and Gingrich,, identification of which party controlled the House before the election, identification of which party controlled the Senate before the election, and identification of the more conservative political party.

4.1. Knowledge as a Dependent Variable

Past studies have identified numerous predictors of political knowledge. Here, I focus on eight: education, age, race, sex, income, interest in politics, political discussion, and internal efficacy. These factors are among those examined by Delli Carpini and Keeter (1996, 182-3), with data drawn from the 1988 NES and their own 1989 Virginia survey. I use these variables because they are the only factors to produce significant effects in both of Delli Carpini and Keeter's data sets. Of these eight, Delli Carpini and Keeter found education and interest in politics to be the most important.[x]

My empirical objective is straightforward. In most past research examining possible predictors of political knowledge, DKs and incorrect answers have been collapsed into a common "absence of knowledge" grouping. Here, I will disaggregate incorrect and DK responses. By doing so, we will be able to determine if these purportedly comparable options do, in fact, respond similarly to the independent variables.

The first step entails estimation of models using the familiar form of the knowledge measures. Two models will be estimated for each year. The first model in each case is estimated using OLS regression; hence, these models resemble those reported in past research. The second model is estimated using grouped-data logistic regression, or a "proportions" model (Greene 1990). In binomial logistic regression, the dependent variable is dichotomous, as it would be here were we estimating whether a person answered a particular knowledge item correctly. The grouped-data model is a variant of logistic regression designed to be used when the underlying distribution is dichotomous, but data have been aggregated across some unit. For instance, we might use a grouped-data model to estimate the district-level Democratic portion of the two-party House vote (the individual-level vote data are coded 1 = Democratic, 0 = Republican; after aggregating by district, the dependent variable would represent the proportion of voters in each district who voted for the Democrat). Here, I aggregate each respondent's answers to the NES knowledge items. Aggregation of these ones (items answered correctly) and zeroes (items not answered correctly) produces proportions data—which are summarized in Table 1.

The grouped-data model enjoys logistic regression's familiar advantages in that predicted proportions will be bounded by zero and one, and the effects of the independent variables need not be

linear. However, the most important advantage of the grouped-data technique for present purposes is its flexibility; the model can be adapted to the multinomial case, an essential virtue if we are to disaggregate incorrect and DK responses.[xi]

Results of the OLS and grouped-data models are shown in Table 2. All 24 OLS coefficients attain statistical significance (ignoring the control for whether the respondent answered the income item), as do virtually all of the corresponding logit coefficients. There is a clear proportionality between the OLS and logit estimates, meaning that the two methods support generally comparable interpretations regarding the relative effects of the eight independent variables. Were analysis to stop here, we would conclude that several consistent correlates of knowledge have been identified, and, viewed from the opposite perspective, that we also understand what factors are related to the absence of knowledge.[xii]

Table 2. Correlates of Correct Response on Knowledge Scales: OLS Regression and Grouped-Data Logistic Regression Estimates

	1992		1994		1996	
	OLS	Logit	OLS	Logit	OLS	Logit
Constant	0.051 (0.025)	-2.113*** (0.277)	0.072* (0.030)	-1.991*** (0.314)	0.256*** (0.039)	-1.234** (0.387)
Education	0.062*** (0.003)	0.290*** (0.034)	0.061*** (0.004)	0.291*** (0.040)	0.042*** (0.004)	0.208*** (0.043)
Age	0.003*** (0.000)	0.014*** (0.003)	0.003*** (0.000)	0.013*** (0.003)	0.001*** (0.000)	0.006# (0.004)
Race	-0.128*** (0.013)	-0.621*** (0.147)	-0.141*** (0.016)	-0.665*** (0.173)	-0.160*** (0.019)	-0.699*** (0.182)
Sex	-0.096*** (0.010)	-0.455*** (0.103)	-0.074*** (0.011)	-0.356** (0.115)	-0.026* (0.013)	-0.131 (0.125)
Respondent answered income item	-0.018 (0.019)	-0.079 (0.204)	-0.010 (0.023)	-0.053 (0.231)	-0.023 (0.028)	0.002 (0.161)
Income	0.003*** (0.001)	0.015 (0.009)	0.004*** (0.001)	0.017# (0.010)	0.005*** (0.001)	0.024* (0.011)
Interest in politics	0.082*** (0.005)	0.380*** (0.058)	0.072*** (0.006)	0.325*** (0.061)	0.067*** (0.007)	0.314*** (0.069)
Political discussion(0.002)	0.011*** (0.024)	0.053* (0.002)	0.016*** (0.025)	0.078** (0.003)	0.006* (0.028)	0.033
Internal efficacy	0.017*** (0.003)	0.082* (0.035)	0.014*** (0.004)	0.069 (0.044)	0.022*** (0.005)	0.105* (0.047)
R^2, χ^2	.496	339.273	.477	209.827	.316	154.760
Number of cases	2,157	2,157	1,704	1,704	1,470	1,470

Source: 1992, 1994, and 1996 National Election Studies.

Note: standard errors are in parentheses. The dependent variable in each model is the proportion of knowledge items answered correctly on nine-item (1992, 1994) or seven-item (1996) knowledge batteries.

*** p < .001 **p < .01 *p < .05 # p < .10

The key test requires that the models be reestimated using grouped-data multinomial logistic regression. Conventional multinomial logistic regression, in which the dependent variable has three or more discrete categories (e.g., the individual-level vote choice in a multi-party election) has seen wide use in political science dating back to Calvert and Ferejohn's (1983) research on coattail voting. The conventional multinomial model could be used to analyze knowledge, but only if responses to each item were considered in separate equations. Nadeau and Niemi (1995) conduct such an analysis,[xiii] and they demonstrate that several predictors produce different effects for incorrect answers and DKs. To determine whether such results replicate when data from knowledge items are summed to form scales, we must abandon conventional multinomial estimation in favor of the grouped-data procedure.

Although I am aware of no previous applications of the grouped-data multinomial technique in political science, this approach provides a direct means to distinguish effects associated with incorrect responses from those associated with DKs. As in standard multinomial logistic regression, where the dependent variable includes k categories, the procedure yields k - 1 columns of estimates. Here, we have three categories: the proportion of items a person answered correctly, the proportion answered DK, and the proportion answered incorrectly. I will use the former as the contrast category, meaning that the first column of results will indicate the effects of the independent variables on the proportion of DK as opposed to correct responses, whereas the second column will contrast incorrect and correct answers.

Coefficients can be compared directly with those in Table 2's binomial models (because correct answers act as the contrast category in the multinomial models, the signs will be the opposite of those in Table 2). The key test concerns whether DK and incorrect responses are comparable. In Table 2, for instance, the logit coefficient for education in 1992 is 0.290; in the multinomial model this variable should yield coefficients of approximately -0.290 for both DKs and incorrect answers. It is too demanding to expect the coefficients to be exactly equal, but three reasonable standards of

comparison can be introduced: 1) coefficients that attain statistical significance for one of the "absence of knowledge" categories also should do so for the other; 2) variance in the magnitude of coefficients should be statistically insignificant; and 3) variance in the magnitude of coefficients should be unsystematic.

The results reported in Table 3 reveal that the assumption that DK and incorrect responses perform similarly misses the mark by a wide margin.[xiv] First, the predictors function much better in differentiating DK from correct responses than in accounting for incorrect answers. Across the three years, 19 of the 24 DK effects reach statistical significance ($p < .05$). In contrast, only eight of the 24 coefficients reach statistical significance for the incorrect vs. correct contrasts. Second, when comparing the 24 pairs of coefficients (superscript letters in the incorrect/correct columns denote statistically significant DK vs. incorrect contrasts), we see that the effects for education and interest in politics—the two most important predictors of knowledge—are significant in all three years, indicating that the propensity to give wrong answers rather than DKs increases as a function of both of these variables. Two additional variables, sex and political discussion, produce marginally-significant ($p < .10$) differences in 1992. Third, and most striking, *variance in coefficient magnitude is fully systematic: in all 24 instances, the DK coefficients are larger than those for incorrect responses*.

Table 3. Disaggregating "Don't Know" and Incorrect Responses: Grouped-Data Multinomial Logistic Regression Results

	1992		1994		1996	
	Don't Know/ Correct	Incorrect/ Correct	Don't Know/ Correct	Incorrect/ Correct	Don't Know/ Correct	Incorrect/ Correct
Constant	2.160*** (0.308)	0.097 (0.390)	1.985*** (0.355)	0.259 (0.426)	1.535*** (0.456)	-0.845 (0.550)
Education	-0.358*** (0.040)	-0.180*** [a] (0.047)	-0.359*** (0.049)	-0.201*** [b] (0.053)	-0.320*** (0.054)	-0.072 [a] (0.057)
Age	-0.015*** (0.003)	-0.013** (0.004)	-0.014*** (0.004)	-0.013** (0.004)	-0.009* (0.004)	-0.003 (0.005)
Race	0.647*** (0.160)	0.569** (0.194)	0.769*** (0.190)	0.493* (0.224)	0.830*** (0.205)	0.474# (0.255)
Sex	0.553*** (0.116)	0.294* [c] (0.142)	0.415** (0.134)	0.294# (0.151)	0.175 (0.152)	0.083 (0.171)
Respondent answered income item	0.077 (0.224)	0.079 (0.293)	-0.027 (0.258)	0.202 (0.324)	-0.106 (0.349)	0.170 (0.435)
Income	-0.019#	-0.005	-0.021#	-0.011	-0.031*	-0.016

	(0.010)	(0.012)	(0.012)	(0.013)	(0.013)	(0.014)
Interest in politics	-0.488*** (0.064)	-0.137# [a] (0.082)	-0.419*** (0.070)	-0.167* [b] (0.081)	-0.433*** (0.082)	-0.137 [b] (0.096)
Political discussion	-0.078** (0.028)	-0.017 [c] (0.032)	-0.100*** (0.030)	-0.051 (0.032)	-0.051 (0.036)	-0.013 (0.038)
Internal efficacy	-0.093* (0.039)	-0.061 (0.049)	-0.080 (0.052)	-0.052 (0.058)	-0.131* (0.056)	-0.068 (0.064)
Number of statistically significant contrasts ($p < .05$)						
DK vs. Correct	7		6		6	
Incorrect vs. Correct	4		4		0	
DK vs. Incorrect	2		2		2	
χ^2	459.724		341.329		201.654	
Number of cases	2,157		1,704		1,470	

Source: 1992, 1994, and 1996 National Election Studies.

Note: standard errors are in parentheses. The dependent variable in each model is the proportion of knowledge items answered correctly, incorrectly, or “don’t know” on nine-item (1992, 1994) or seven-item (1996) knowledge batteries. The contrast category is the proportion answered correctly.

*** $p < .001$ **$p < .01$ *$p < .05$ # $p < .10$

Superscript letters indicate statistically significant effects for the don’t know/incorrect paired contrast; these contrasts function as tests of the equality of the coefficients for the don’t know/correct and incorrect/correct contrasts: a) $p < .001$, b) $p < .01$, c) $p < .10$

Multinomial logit coefficients are difficult to interpret substantively, and thus graphical display of results often yields additional insight. Figure 2 depicts predicted proportions for each independent variable; in each panel, the coefficients for the other predictors are held at their means or modes as appropriate. In viewing these results, it is important to keep in mind that the graphs depict simulated values, and that there is uncertainty surrounding these estimates (i.e., the models include a stochastic element). In examining Figure 2, it is the general pattern of results, not the precise values of specific estimates, that is of interest. The general pattern of results is clear. In nearly every instance, the solid black lines, which indicate the predicted proportions of incorrect answers, are essentially *flat*, and what little slope there is to these lines runs in the incorrect direction as often as not. In all years, people with the highest levels of education answer as many questions wrongly as do people with the lowest levels of education. Men get as many wrong as women. People who are interested in politics fare no better than people who are not. And so on. Even those few incorrect vs. correct effects that are statistically significant turn out to be unimpressive: the "slopes"

for incorrect answers generally differ from those for correct answers and DKs only when movement between the latter two is so extreme as to cause the side-effect of movement relative to what is nothing more than a stationary object.[xv]

Patterns in Figure 2 resemble those in panel B of Figure 1. Current knowledge scales build on the intuition that right is the opposite of wrong. Empirically, however, correct and incorrect answers do not behave as opposites, and DKs and incorrect answers do not perform comparably. What do these findings imply regarding whether knowledge is discrete, and for the possibility that our scales tap constructs other than knowledge? Clearly, the results can be read as consistent with either or both scenarios, but not with the view that current scales tap a single, discrete construct. Concerning whether knowledge is discrete, it may be that the trade-off between correct answers and DKs reflects movement from the complete failure to receive and understand information to high marks on both reception and comprehension. For instance, respondents with low levels of education and interest in politics say "don't know" quite a lot, but they also tend to answer one or two items incorrectly. This pattern may mean that such respondents receive very little information, and what they do receive they misunderstand. In contrast, respondents with high levels of education and interest in politics receive a great deal of information, and thus they rarely say "I don't know." Their understanding of information is good, but not perfect, as indicated by the fact that they tend to answer one or two items incorrectly. Perhaps a little misunderstanding is the best that we can expect from some people, and the worst that we can expect from others.

Current results also may indicate that knowledge scales measure respondents' personality traits. When a person answers an item incorrectly, we are certain that the person is not fully informed. In contrast, respondents can answer questions correctly with little or no information, and they can say DK even when they are partially or fully informed. The category that is most resistant to personality effects (incorrect answers) refuses to budge in response to the independent variables. Thus, a case could be made that current results are more consistent with the claim that our scales represent personality attributes than with the claim that they measure knowledge. At minimum, personality factors may confound knowledge scales. Some DKs may reflect a response bias in which people lacking in self-confidence say "don't know" out of fear of answering incorrectly. If these individuals could be induced to offer substantive replies, many of those answers likely would be

correct. From this perspective, personality-based confounds possibly have led us to overstate both variance in levels of political knowledge and our capacity to explain that variance (if the "correct answer" lines in Figure 2 reflect a mix of knowledge and personality, then they would be less steep if they tapped only knowledge).

4.2. Knowledge as an Independent Variable

Scholarly concern with political knowledge constitutes more than a simple descriptive exercise. Knowledge is thought to influence how much and how well citizens participate in politics. Thus, understanding knowledge is vital if we are to heighten participation and improve representation. As an empirical matter, it follows that measures of political knowledge can be used as predictors in models examining a wide array of behavioral and attitudinal variables. And, given the curious behavior of knowledge measures observed thus far, it also follows that the influence of knowledge on our dependent variables may be misspecified if we fail to disaggregate DK and incorrect responses. To demonstrate this point, I focus on voter turnout, a dependent variable that has been examined in two prominent works on political knowledge and political sophistication (Delli Carpini and Keeter 1996; Neuman 1986).

The approach here is quite simple. The dependent variable is self-reported voter turnout (1 = respondent reported having voted, 0 = respondent reported having not voted), with data again drawn from the 1992, 1994, and 1996 NES surveys.[xvi] Independent variables include the eight factors used above, along with two specifications of knowledge itself. The first model for each year operationalizes knowledge using only one variable, the proportion of items not answered correctly (i.e., the sum of the DK and incorrect answer proportions). The second model replaces this summary measure with separate indicators for DKs and incorrect answers. Once again, if “don’t know” means the same thing as a wrong answer, then the two variables should perform similarly. As predictors of turnout, the two slopes should be of comparable magnitude.

Logistic regression results are reported in Table 4. In all years, the DK coefficient exceeds the incorrect answer coefficient in both magnitude and significance level. The disparity is minor in 1996, moderate in 1992, and quite sizable in 1994.[xvii] Collectively, the effects in Table 4 reveal that the form of incorrect response matters: respondents are more likely to vote if the knowledge items they did not answer correctly were answered incorrectly rather than DK. One form of ignorance is

associated with higher voter turnout than is the other. This effect is displayed in Figure 3. The dotted lines represent a scenario in which all questions not answered correctly were answered with DKs, whereas the solid lines indicate that all items not answered correctly were answered incorrectly (no respondent answered more than six items incorrectly; the solid lines give way to dashed lines to represent this point).

Table 4. Responses on Knowledge Scales and Self-Reported Voter Turnout: Logistic Regression Results

	1992		1994		1996	
	Model I	Model II	Model I	Model II	Model I	Model II
Constant	-0.728# (0.442)	-0.718 (0.443)	-1.255** (0.434)	-1.234** (0.437)	-0.786 (0.599)	-0.772*** (0.601)
Education	0.317*** (0.053)	0.309*** (0.053)	0.131** (0.048)	0.123* (0.048)	0.201*** (0.058)	0.199*** (0.059)
Age	0.021*** (0.004)	0.022*** (0.004)	0.029*** (0.004)	0.030*** (0.004)	0.026*** (0.005)	0.026*** (0.005)
Race	0.158 (0.167)	0.160 (0.168)	0.135 (0.187)	0.176 (0.189)	-0.035 (0.220)	-0.041 (0.221)
Sex	0.551*** (0.135)	0.564*** (0.135)	0.384** (0.131)	0.395** (0.132)	0.524*** (0.159)	0.524*** (0.159)
Respondent answered income item	-0.465# (0.249)	-0.475# (0.251)	-0.824** (0.271)	-0.874** (0.273)	-1.385** (0.460)	-1.391** (0.461)
Income	0.034** (0.011)	0.033** (0.012)	0.021# (0.011)	0.020# (0.011)	0.048*** (0.013)	0.047*** (0.013)
Interest in politics	0.375*** (0.071)	0.350*** (0.072)	0.371*** (0.069)	0.340*** (0.070)	0.472*** (0.087)	0.469*** (0.088)
Political discussion(0.034)	0.110** (0.034)	0.105** (0.029)	0.151*** (0.029)	0.149*** (0.043)	0.211*** (0.043)	0.211***
Internal efficacy	0.157*** (0.043)	0.156*** (0.043)	0.194*** (0.050)	0.193*** (0.050)	0.216*** (0.060)	0.215*** (0.060)
Knowledge items not answered correctly (sum of DK and incorrect)	-2.083*** (0.284)		-2.054*** (0.274)		-1.572*** (0.295)	
"Don't Know"		-2.227*** (0.295)		-2.538*** (0.325)		-1.611*** (0.325)
Incorrect		-1.405** (0.464)		-0.881# (0.462)		-1.457** (0.491)
χ^2	577.762	581.235	553.845	563.949	360.365	360.451
Number of cases	2,156	2,156	1,703	1,703	1,470	1,470

Source: 1992, 1994, and 1996 National Election Studies.

Note: standard errors are in parentheses. The dependent variable in each model is self-reported voter turnout: 1 = respondent reported voting, 0 = respondent did not report voting.

*** p < .001 **p < .01 *p < .05 # p < .10

Results for 1992, and especially 1994, justify concern that DKs and incorrect answers function differently from one another as independent variables. Results for 1996 follow the same basic pattern as the two previous elections, but the effect is substantively negligible in 1996. Regardless of what we make of 1996, results across the three elections are sufficiently consistent to warrant additional concern about the use of conventional measures of political knowledge. Once again, DKs and incorrect answers have performed dissimilarly, impeaching the common practice of collapsing these two categories.

Current results comport well with the possibility that knowledge is not discrete. If acquisition of information motivates one to vote, then it may not be essential that that information be accurate. Citizens who think that they are well-informed even when they actually are misinformed perhaps demonstrate a psychological state akin to that of the genuinely well-informed. The misinformed don't know that they don't know. Obviously, severe misinformation (thinking that voter registration has been eliminated; believing that the election will be held in February this time; thinking that you can send your child to vote for you) can limit turnout, implying that turnout should be highest when information has been both received and comprehended.

Personality-based factors also may drive the effects reported in Table 4. If DKs reflect a lack of self-confidence rather than the true absence of information, then low turnout among respondents with high DK rates may reveal unwarranted reticence. Some people may defer from voting because they feel that they know too little about politics, yet other people with the same level of information do vote.

5. Is There a Personality Effect?

Results reported thus far are consistent with both the possibility that knowledge is not discrete and the possibility that knowledge scales are contaminated by personality. The first of these concerns is less serious than the second. If knowledge scales do not tap personality, then use of these data will be improved simply by disaggregating incorrect answers and DKs. Survey questions would not need to be revised; only our statistical approaches would have to change. Matters become more complicated if a personality confound exists. In educational testing, the risk of personality-based

contamination is limited through exclusion of DK as a choice option. Implementation of this solution in social science would require a new protocol on our surveys; we would have to begin discouraging rather than encouraging DKs. Clearly, it is important that we determine whether personality does, in fact, contaminate current knowledge measures.

This matter will not be resolved conclusively here, but all available evidence suggests that the threat of a personality effect is considerable. Four points warrant mention. First, there is no a priori reason to believe that the standard result from the field of educational testing and psychology—that a DK option invites personality-based contamination—will not be found with survey research. To the contrary, the logic outlined in research on educational testing applies even more powerfully to the opinion survey. On educational tests, students are motivated to do their best, which may attenuate the impact of personality. On surveys, respondents gain nothing by performing well on knowledge batteries, meaning that no external incentives limit the possible effects of personality. For respondents lacking in self-confidence, surveys provide no reason *not* to seek shelter in the DK response.

Second, analysis of NES data provides circumstantial evidence consistent with the claim that knowledge scales tap something in addition to knowledge. We saw previously that a respondent's lack of interest in the survey corresponds with a higher DK rate, and thus with a lower knowledge score. The list of standard predictors of knowledge also arouses suspicion. The sex gap on standardized tests has been shown to be related to differences between men and women in self-confidence, and the same may be true for the gap in knowledge about politics. We also might reasonably expect that self-confidence will affect claimed levels of interest in politics, efficacy, and political discussion.

Third, a split-ballot experiment provides direct evidence regarding the influence of personality.[xviii] Results should be viewed as preliminary in that the participants were university undergraduates and the sample was small. The instrument included 21 knowledge items. The 94 participants were assigned randomly to one of two treatments. The first condition used multiple-choice items, respondents were encouraged to answer all items, and no DK option was included. The second condition also used multiple-choice items, but respondents were discouraged from guessing, and a DK option was included with each item. Consistent with current recommendations in educational testing (Haladyna and Downing 1993; Landrum, Cashin and Theis 1993), all multiple-

choice items included three choice options.

A knowledge scale was created using data from 15 of the knowledge items.[xix] Results for the individual items and for the aggregate scale are reported in Table 5. Comparing columns one and two, we see that the correct response rate of participants who were discouraged from answering DK equaled or exceeded that of the second group on all 15 items, and that the difference between the two aggregate measures is statistically significant.

Table 5. The Effect of Encouraging and Discouraging DKs on Knowledge Data
(column entries are the percentage of participants answering correctly)

Item	Cell One : DKs Discouraged (n = 47)	Cell Two: DKs Encouraged (n =47)
Whose responsibility is it to determine if a law is constitutional or not?	76.6	68.1
Would you say that one of the parties is more conservative than the other at the national level?	78.7	74.5
What job or political office is now held by Newt Gingrich?	95.7	87.2
What job or political office is now held by William Rehnquist?	57.4	31.9
What job or political office is now held by Trent Lott?	38.3	21.3
What job or political office is now held by Al Gore?	95.7	95.7
What job or political office is now held by Boris Yeltsin?	93.6	76.6
Whose responsibility is it to appoint federal judges and justices?	76.6	61.7
How long is a U.S. Senate term?	44.7	40.4
How many members are there in the U.S. House?	63.8	31.9
How many members are there in the U.S. Senate?	70.2	51.1
How long is a term on the U.S. Supreme Court?	87.2	87.2
Can a communist run for president in the United States?	44.7	31.9
Which of the following is a U.S. Senator from Florida?	87.2	70.2
How old must a person be to be elected to the U.S. House?	63.8	25.5
Average ($t = 3.30$, $p < .01$)	71.6	57.0

Note: Data are from 94 university undergraduates. All items used a three-category multiple-choice format. Participants in cell one were discouraged from answering DK, and no DK option was provided; participants in cell two were encouraged to answer DK if they were not certain of an answer, and a DK option was included with each item.

Three pieces of evidence suggest that personality affected results when participants were encouraged to choose DK. First, if personality played no role, then differences between the two cells of the experiment should result solely from blind guessing by participants who were discouraged from choosing DK. We can test for the effect of blind guessing by awarding participants partial credit for each DK (i.e., substituting a blind guess for a DK would bring a one-in-three chance of a correct

answer). This modification increases the average correct answer rate for cell two from 57.0 to 66.0, implying that a majority of the gap shown in Table 5 is the consequence of blind guessing. However, nearly 40 percent of the gap persists. This suggests that some partially-informed respondents chose DK in cell two of the experiment while their counterparts in cell one offered correct answers.

Second, following Cronbach's (1946) claim that personality-based confounds artificially inflate reliability, we can compare the inter-item correlations for the knowledge scales produced in the two cells of the experiment. The average correlation for cell one is .16, compared with .23 for cell two. In the cell that is at risk of a personality effect, Cronbach's prediction of inflated reliability is supported.

Third, the study included a battery of personality items, meaning that the possible link between personality and knowledge can be tested directly. A seven-item self-confidence scale was formed. The knowledge score for cell one was not found to be related to self-confidence ($t = 0.44$, $R^2 = .00$), but an effect was found for cell two ($t = 3.21$, $R^2 = .19$). Several competitiveness and risk-taking items also were included on the personality battery, but most had no effect on either knowledge measure. One exception was an item concerning whether the respondent would make a high-risk investment; when this item was included along with self-confidence in the regression models, there again were no effects for cell one (self-confidence, $t = 0.38$; risk-taking, $t = 0.05$; $R^2 = .00$), but both self-confidence and risk effects appeared for cell two (self-confidence, $t = 2.67$; risk-taking, $t = 2.12$; $R^2 = .26$).

Although these are only preliminary data, all signs point to a strong personality effect. Well-specified knowledge models such as those reported in Table 2 produce R^2 values of approximately .50. Two personality variables team to perform half as well here. Coupled with evidence that the DK option deterred more than blind guessing and the evidence of inflated reliability, the direct personality effect identified here indicates that current measurement practices may invite personality-based contamination.

Past NES surveys provide additional evidence that personality may influence responses on knowledge items. The 1956 and 1958 NES surveys included questions similar in content to conventional measures of self-confidence and risk-taking. A self-confidence scale was created using data from three dichotomous items.[xx] A fourth dichotomous item is used to measure risk-taking.[xxi]

Unfortunately, the 1956 NES included no knowledge items, and the 1958 survey included only the item regarding which party controlled the House prior to the election. Using this item as the dependent variable, the effects of personality can be explored. Results are reported in Table 6. In Model I, binomial logistic regression estimates reveal that the propensity to answer the party-control item correctly increases as a function of both self-confidence and risk-taking. In Model II, DKs are used as the contrast category in a multinomial specification. Here, we see that correct answers and incorrect answers both increased relative to DKs as a function of the two personality variables. Respondents who claimed to be sure of themselves and open to new things tended to provide substantive answers—both correct and incorrect—on the party control of the House item, whereas less secure, risk-averse respondents were more likely to say DK.

The possibility that knowledge measures are contaminated by personality has been considered from four perspectives. First, the logic underlying research in educational testing applies as well or better to survey research, implying that we should not presume political knowledge measures to be free of personality effects. Second, data from the 1992-96 NES surveys suggest that knowledge scales tap something other than knowledge: DK rates increase as a function of lack of interest in the survey, and DK rates vary with familiar correlates of personality such as sex. Third, data from a split-ballot study reveal a direct link between personality and knowledge for participants who were encouraged to answer DK, and no such link was detected for participants who were discouraged from answering DK. Fourth, indicators of self-confidence and risk-taking formed with data from the 1958 NES were found to be related to the propensity to answer DK on a standard knowledge item. Although none of these tests by itself offers conclusive proof that knowledge scales suffer personality-based contamination, the evidence is of sufficient collective strength that remedial action should be considered.

6. Implications and Recommendations

The construction of well-behaved measures constitutes one of the key challenges facing researchers in the social sciences. The central conclusion to be drawn from this paper's discussion and analysis is that, for political knowledge, this challenge still remains. Viewed empirically, the dissimilar performance of "don't know" and incorrect answers—response categories defined in our knowledge scales as indicators of an identical behavioral state—signals that something is amiss. Factors

previously thought to affect levels of political knowledge have been shown to influence DK responses, yet these same factors exert virtually no effect on the number of knowledge items respondents answer incorrectly. Similarly, voter turnout has been shown to decline sharply as DKs rise, but to vary only moderately as a function of incorrect answers. Conventional knowledge measures are anything but well-behaved.

Two threats to knowledge scales have been considered. These threats have been assessed separately, but it is important to note that the possibility that knowledge is not discrete is not mutually exclusive with the possibility that our scales are contaminated by personality. Final word on personality effects must await survey experiments with subjects drawn from the general population. Barring the emergence of exculpatory evidence, the most prudent course is to assume that both of the problems discussed here do, in fact, affect current scales. If this is the case—if knowledge is not discrete, and if a DK option brings contamination—then what can be done to rectify these problems?

Measuring Political Knowledge. For researchers designing measures of political knowledge for future surveys, the risk of personality-based contamination demands consideration. Available evidence indicates that surveys that encourage DKs invite such contamination, whereas the protocol of including a strong DK option was adopted without adequate attention to this risk. Indeed, one reason offered in defense of the current approach is that encouraging DKs increases scale reliability—even though the research of Cronbach and others demonstrated that this boost in reliability occurs precisely because personality contaminates the scales.

If we presume that DKs invite personality effects, then the obvious recommendation is that future surveys should discourage DKs. Introductions to knowledge batteries should emphasize to respondents that it is important that they try to answer each question, the items should not include DK options, and interviewers should prompt respondents for substantive answers. When DKs are provided, they should be assigned randomly to the available choice options to account for the effects of blind guessing. Additionally, it would be useful if more split-ballot data were available so that the consequences of encouraging and discouraging DKs could be assessed in head-to-head tests.

We cannot infer greater sophistication from incorrect answers than from DKs if DK no longer exists as a choice option. Thus, an alternate means to distinguish fully-informed, partially-informed, and uninformed respondents is required. One approach is to include knowledge items that

are ordinal in form. For instance, ordinal data would be produced by a question that asked respondents to name the freedoms protected by the First Amendment to the Constitution. The same end would be achieved if knowledge batteries included a sufficient number of items for the fully informed, the partially informed, and the uninformed to sort themselves out (much as the A, B, C and D students sort themselves out on academic examinations). Delli Carpini and Keeter (1993, 1996) offer a five-item knowledge battery that achieves a reasonable level of reliability (alpha = .71). If reliability is inflated due to personality-based contamination, future knowledge batteries will need to include more than five items to achieve acceptable reliability. For example, if personality accounts for one-third of the correlation among Delli Carpini and Keeter's five items, then a scale free of personality effects would need nine items to produce an alpha of .71.[xxii] Just as few of us would feel comfortable measuring our students' knowledge levels with a handful of multiple-choice and short-answer items, we must recognize that political knowledge may not be represented adequately if our scales are too brief.

A final matter in the measurement of political knowledge is what response format our items should use. NES items include two-category multiple-choice questions, three-category multiple-choice questions, and open-ended identification questions. The advantages of multiple-choice items have led analysts in educational testing to voice support for this format, particularly when three response options are provided. As noted above, social scientists have not exploited a key advantage of multiple-choice items, the capacity to measure conceptual material. Hence, in designing future knowledge batteries, we should strive to tackle topics that require more than the recall of basic facts. Short-answer, or identification, items have seen frequent use in knowledge batteries. The short-answer format produces a large number of DKs, bringing the risk that partial information goes undetected. Thus, if open-ended identification items continue to be used, it is important that both the wording of the questions and the behavior of survey interviewers encourage substantive response.

Knowledge as a Dependent Variable. For researchers designing their own surveys, the absence of a DK option eliminates the risk of personality-based contamination. Unfortunately, no comparable solution exists for analysts examining extant data. If past surveys have tapped personality, there is no post hoc means to rectify this problem. When knowledge is used as a dependent variable, indicators of relevant personality attributes could be included as

predictors—except, of course, that the requisite personality measures appear on virtually no surveys in political science.

Although the problems discussed in this paper cannot be cured in research using on existing data, it is possible to treat the symptoms. The common practice of grouping incorrect answers and DKs must be discontinued. If the dependent variable under consideration draws on data from a single knowledge item, standard multinomial estimation techniques should be employed to disaggregate DK and incorrect responses. Nadeau and Niemi (1995) distinguished DK and incorrect responses through use of multinomial logistic regression, and found that several predictors produced differing effects for these response options. Future research should follow this lead. Likewise, when knowledge scales are used as dependent variables, grouped-data estimation procedures such as the grouped-data logistic regression technique used in this paper are appropriate.

Nadeau and Niemi's (1995) results are consistent with the notion that knowledge items tap an ordinal construct: incorrect responses may fall mid-way between correct answers and DKs because partial information reflects greater sophistication than does the lack of information. Because this ordinal pattern also is consistent with personality-based effects, it is premature to conclude that knowledge questions capture a single ordinal construct. However, this possibility also should not be dismissed, and thus use of statistical techniques designed for ordinal data should be considered. The bottom line is that while DK and incorrect responses should not be collapsed, it has not yet been resolved whether it is most appropriate to treat the three response options on knowledge items as ordered or discrete selections.

Knowledge as an Independent Variable. Personality-based contamination brings two risks when knowledge is used as a predictor. In cases such as the example of the effect of knowledge on voter turnout considered here, it is possible that past research has mislabeled what is in part a personality effect. In other cases, though, it seems unlikely that it is personality rather than knowledge at work. Most research using knowledge as an indicator of political sophistication fits this latter category. In these cases, personality, like other sources of error in our scales, would bring a conservative effect. Hence, the emergence of knowledge data free from personality-based contamination should lead to strengthened results in most situations in which knowledge is used as an independent (or moderating) variable.

If current knowledge scales are ordinal, then collapsing DK and incorrect answers typically will merge strong and weak effects, costing analytical precision. Here, the solution is simple. Just as we use often use two dummy variables to distinguish three categories of partisanship (Democrat, independent, Republican) and ideology (liberal, moderate, conservative), we must use two variables to distinguish the three options on our knowledge scales. Where DKs and incorrect responses perform dissimilarly, disaggregation will add explanatory power.

The complexity of human behavior means that it is impossible for measurement in the social sciences to gain the precision characteristic of measurement in other fields. Nonetheless, we must strive to develop indicators of the highest possible quality. Given the centrality of political knowledge in research on mass political behavior, it is especially important that knowledge be measured well. This paper has demonstrated that improvements are possible on several levels. In designing and using knowledge scales, we must heed the lessons learned in over 70 years of research by scholars in the field of educational testing and psychology, and we must resist the temptation to assume that different choice options—incorrect answers and DKs—mean the same thing. Attention to the concerns raised in this paper should improve our capacity to explore the antecedents and consequences of political knowledge.

References

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

Althaus, Scott. L. 1998. “Information Effects in Collective Preferences.” *American Political Science Review* 92:545-58.

Bartels, Larry M. 1996. “Uninformed Votes: Information Effects in Presidential Elections.” *American Journal of Political Science* 40:194-230.

Beidleman, David C., and Christine L. Cole. 1991. "Scholastic Aptitude Test Gender Gap." *American Secondary Education* 19(no.2):2-5.

Bennett, Linda L.M., and Stephen Earl Bennett. 1989. "Enduring Gender Differences in Political Interest.: The Impact of Socialization and Political Dispositions." *American Politics Quarterly* 17:105-22.

Bennett, Stephen Earl. 1989. "Trends in Americans' Political Information, 1967-1987." *American Politics Quarterly* 17:422-35.

Bennett, Stephen Earl. 1988. "'Know-Nothings' Revisited: The Meaning of Political Ignorance Today." *Social Science Quarterly* 69:476-90.

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

Calvert, Randall L., and John A. Ferejohn. 1983. “Coattail Voting in Recent Presidential Elections.” *American Political Science Review* 77:407-19.

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." *Developmental Psychology* 33:669-80.

Cassel, Carol A., and Celia C. Lo. 1997. "Theories of Political Literacy." *Political Behavior* 19:317-35.

Converse, Philip E. 1964. "The Nature of Belief Systems in Mass Publics." in *Ideology and Discontent*, ed. David E. Apter. New York: Free 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. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

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

Francis, Joe D., and Lawrence Busch. 1975. "What We Know About 'I Don't Knows.'" *Public Opinion Quarterly* 39:207-18.

Goren, Paul. 1997. "Political Expertise and Issue Voting in Presidential Elections." *Political Research*

Quarterly 50:387-412.

Graber, Doris A. 1996. “Wrong Questions, Wrong Answers: Measuring Political Knowledge.” Paper presented at the annual meeting of the Midwest Political Science Association, Chicago.

Greene, William H. 1995. *Limdep Version 7.0 User’s Manual.* Plainview, N.Y.: Econometric Software.

Greene, William. H. 1990. *Econometric Analysis.* New York: Macmillan Publishing Company.

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

Hahn, Carole L. 1996. "Gender and Political Learning." *Theory and Research in Social Education* 24:8-35.

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.

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

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.

Iyengar, Shanto. 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. Urbana: University of Illinois Press.

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

Kuklinski, James. H., Paul J. Quirk, David W. Schwieder, and Robert F. Rich. 1998. “’Just the Facts, Ma’am’: Political Facts and Public Opinion.” *The Annals of the American Academy of Political and Social Science* 560:143-54.

Lambert, Ronald D., James E. Curtis, Barry J. Kay, and Steven D. Brown. 1988. "The Social Sources of Political Knowledge." *Canadian Journal of Political Science* 21:359-74.

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.

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

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. 1995. "Newspapers and Political Awareness." *American Journal of Political Science* 39:513-27.

Nadeau, Richard, and Richard G. Niemi. 1995. “Educated Guesses: The Process of Answering Factual Knowledge Questions in Surveys.” *Public Opinion Quarterly* 59:323-46.

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

Neuman, W. Russell, Marion R. Just, and Ann N. Crigler. 1992. *Common Knowledge: News and the*

Construction of Political Meaning. Chicago: University of Chicago Press.

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

Pettey, Gary R. 1988. “The Interaction of the Individual’s Social Environment, Attention and Interest,

and Public Affairs Media Use on Political Knowledge Holding.” *Communication Research* 15:265-81.

Price, Vincent, and John Zaller. 1993. "Who Gets the News: Alternative Measures of News Reception

and Their Implicatons for Research." *Public Opinion Quarterly* 57:133-64.

Rapoport, Ronald B. 1985. "Like Mother, Like Daughter: Intergenerational Transmission of DK Response Rates." *Public Opinion Quarterly* 49:198-208.

Rapoport, Ronald B. 1982. "Sex Differences in Attitude Expression: A Generational Explanation." *Public Opinion Quarterly* 46:86-96.

Rapoport, Ronald B. 1979. "What They Don't Know Can Hurt You." *American Journal of Political Science* 23:805-15.

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.

Stanley, Jullian C., and Kenneth D. Hopkins. 1972. *Educational and Psychological Measurement and*

Evaluation. Englewood Cliffs, N.J.: Prentice-Hall.

Verba, Sidney, Nancy Burns, and Kay Lehman Schlozman. 1997. “Knowing and Caring About Politics: Gender and Political Engagement.” *Journal of Politics* 59:1051-72.

Weaver, David, and Dan Drew. 1993. "Voter Learning in the 1990 Off-Year Election: Did the Media Matter?" *Journalism Quarterly* 70:356-68.

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.

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

Notes

[i]. For a discussion of the implications of misinformation, see Kuklinski et al. 1998.

ii. It is important to distinguish between two types of guessing. "Guessing" implies blind guessing, or random behavior. However, respondents who are partially informed offer an informed or educated guess (Mehrens and Lehmann 1984). Mehrens and Lehmann argue that informed guessing is more common on educational tests than is blind guessing. This also may be the case with measures of political knowledge (when respondents misidentify William Rehnquist, few say that he is the shortstop for the Mets or the original drummer for the Beatles). Corrections can be introduced to account for guessing, but guessing penalties have been widely discredited in the educational testing literature. Guessing penalties do not discriminate between the blind guess, the informed guess, and the genuine misinformed response. As a result, they tend to do more harm than good, particularly when only a few items appear on the test.

iii. A second benefit of the preface suggested by Delli Carpini and Keeter (1993, 1996) is that it makes the knowledge measure less awkward for those respondents who feel uncomfortable about their lack of political awareness. This is an important practical consideration. In my own surveys measuring political knowledge, I monitored many of the interviews. It was clear that this portion of the survey often was demoralizing for those respondents who fared poorly. Many respondents seemed embarrassed despite the heroic efforts of the interviewers to counter this effect. Hence, if wording that discourages guessing confounds the measurement of knowledge (and I believe it likely does), one practical issue we will face is how to improve knowledge measures while avoiding damage to the psyches of our respondents.

[iv]. NES data allow a test of the suggestion that DKs partly reflect boredom. Interviewers rate respondents on several dimensions, including level of information, intelligence, and interest in the interview. If DKs represent a lack of knowledge, and if interviewers are accurate in their appraisals, then DKs should be most common among respondents who interviewers rate as low in information and intelligence; the same relationships should hold for incorrect answers. However, if some people choose DK because they are too disinterested in the interview to provide substantive answers, this

may be tapped by interviewers' perceptions of respondents' interest levels; this effect should not hold for incorrect answers. A problem in estimating these relationships is that interviewers' ratings may be affected by how respondents answered the knowledge items; for instance, the fact that a respondent repeatedly selected DK on the knowledge items may have led the interviewer to rate the respondent as low in information, intelligence, and perhaps interest. To avoid this problem, I used interviewer ratings from the 1996 NES preelection survey as predictors (six of the seven knowledge items were asked on the postelection survey). The dependent variables were, first, correct answers / (correct + DK), and second, correct answers / (correct + incorrect). Ratings of information and intelligence produced the expected positive effects in both models; ratings of respondents as informed and intelligent corresponded with the tendency of those respondents to give correct answers rather than to say DK, and to give correct answers rather than incorrect answers. In the DK model, the interest variable also produced a positive effect ($t = 2.16$), meaning that the tendency to select DK is inversely related to interest, controlling for the effects of information and intelligence. Conversely, interest produced a negative coefficient ($t = 2.13$) in the incorrect answer model, implying that interest in the interview compels respondents with low levels of information and intelligence to offer substantive responses to knowledge items. A final test used the dependent variable incorrect / (incorrect + DK). The interest measure produced an effect ($t = 3.08$), with DKs again found to be inversely related to interest. These findings are consistent with the claim that some respondents may choose DK for reasons unrelated to knowledge. Likewise, some respondents may guess, and sometimes guess correctly, simply because they are interested in the survey interview.

v. Guessing is discouraged on tests such as the SAT and the GRE, and these exams impose guessing penalties. Controversy surrounding these exams (along with watching re-runs of "The Family Feud") prompted me to reconsider the measurement of political knowledge. A persistent finding in the literature on political knowledge is that men know more about politics than women (Bennett and Bennett 1989; Delli Carpini and Keeter 1996; Lambert et al. 1988; Verba, Burns and Schlozman 1997; cf., Hahn 1996). Delli Carpini and Keeter (1996) note that only a portion of this gap can be accounted for by gender differences in socioeconomic status, interest in politics, etc. Similarly, a somewhat mysterious gender gap is regularly reported on standardized tests, with males receiving

higher scores. Given that this gap does not correspond with a performance gap among undergraduates and graduate students, critics have argued that these tests partly measure differences in personality or socialization rather than only achievement. Because women tend to leave more items blank than do men, and because men tend to be more competitive and self-confident than women, the personality/socialization hypothesis seems plausible. Consistent with this view, recent research has found that the gender gap on a GRE subject test and on the SAT stem partly from variance in levels of self-confidence (Casey, Nuttall and Pezaris 1997; Hirschfeld, Moore and Brown 1995; cf., Beidleman and Cole 1991).

vi. I have not yet discussed the efforts of various researchers, but especially Delli Carpini and Keeter (1993, 1996), to identify which items should be used to form a general scale of political awareness. In such endeavors, the central means to evaluate particular items are to determine which items are well correlated with one another and with the composite scale, and which are correlated with criterion variables. These evaluations must be reconsidered if it becomes clear that knowledge scales partly reflect respondents' personality traits. For example, Delli Carpini and Keeter draw on item-response theory to inform their tests of scale construction. However, item-response theory assumes that only a single construct is being measured, or, at worst, that the construct under consideration is the dominant factor affecting test performance (Hambleton and Swaminathan 1985). Given that some analysts have claimed that objective tests of the form used to measure political knowledge identify personality traits more reliably than they measure achievement, it may be the case that knowledge is not the dominant factor affecting test performance. Put another way, knowledge items that perform well may do so partly because they measure a common personality trait such as self-confidence.

vii. Because the identification items use a short-answer format, some mention of the features of such items is warranted. Problems may exist with multiple-choice items in that respondents with partial information can answer correctly or incorrectly (i.e., knowledge is not discrete), respondents with no information can answer correctly (the correct blind guess), and a DK option may invite personality-based contamination. Short-answer questions are immune to the second of these problems, blind guessing. However, a short-answer format does not preclude the educated guess (a person who is asked

to identify William Rehnquist might associate Rehnquist with President Clinton's impeachment trial, and thus know that Rehnquist is either a senator or a Supreme Court justice, yet be uncertain of which). Additionally, the short-answer format may exacerbate personality-based confounds (the shy respondent who thinks that Rehnquist is the Chief Justice of the Supreme Court will be reassured upon hearing that position listed among a multiple-choice item's options). Two additional limitations are unique to short-answer items. First, the scoring of these items is subjective. We impose discrete scoring (answers are either right or wrong), even though some wrong answers are "less wrong" than others. Is it correct or incorrect when the respondent says that Rehnquist is "a judge"? Second, the short-answer format restricts content to straightforward factual material; multiple-choice items are superior if the examiner seeks to tap more conceptual subject matter (see Aiken 1986; Ebel and Frisbie 1986; Mehrens and Lehmann 1984; Stanley and Hopkins 1972). Unfortunately, political scientists have not capitalized on this potential strength of multiple choice items. Questions used to measure political knowledge focus mostly on basic facts, with little attention to underlying concepts, a concern discussed by Neuman, Just and Crigler (1992) and Graber (1996).

[viii]. Delli Carpini and Keeter's fifth item, which concerns the size of a majority needed to override a presidential veto, was not asked by the NES in the years considered here.

ix. Historical differences between 1996 and the earlier years worked to the advantage of respondents in 1996: 1) a great deal of media attention was paid to the Republican victories in the House and Senate in 1994; 2) vast attention also was given to Newt Gingrich as Speaker; 3) together, these first factors contributed to the heightened salience of ideology; and 4) as a sheer function of time, respondents had additional opportunity to learn the identities of Boris Yeltsin and William Rehnquist.

x. The variables are coded as follows: education 0 to 6; age = age in years; race = 1 if black, 0 if other; sex 1 = female, 0 = male; income = 0 to 23; interest in politics = 0 (low) to 3; political discussion = number of days in the past week the respondent discussed politics; efficacy = 0 (low) to 4. For income, respondents who declined to answer this item are coded as 0, and a separate dummy variable is used to differentiate respondents who did and did not answer the income item (1 =

respondent answered the item); this procedure has no effect on results for income, but substantially increases the number of cases included in the models.

Like Delli Carpini and Keeter (1996), I make no strong assumptions regarding the nature of the relationships between the independent variables and knowledge. For instance, it may be the case that political knowledge heightens one's interest in politics rather than interest leading to knowledge.

xi. Because conventional knowledge measures report count data (i.e., the scales report how many items were answered correctly), Poisson regression might have been appropriate. However, as indicated in table 1, the data more closely resemble a normal distribution than a Poisson distribution. In any case, the grouped-data approach is preferable in that it allows for multinomial estimation, as reported below.

[xii]. Caution must be exercised in some applications of grouped-data logistic regression. Greene (1995) notes that weighting should be used when the number of observations aggregated to form a data point varies across cases, such as if we were to aggregate vote data by state. However, a consequence of weighting is that standard errors can be implausibly small. In the current application, weighting is unnecessary because the same number of observations is aggregated in each case, and thus standard errors are not deflated. To the contrary, when viewed relative to their corresponding coefficients, standard errors in table 2 are lower in every case for the grouped-data models than for the OLS models.

[xiii]. Nadeau and Niemi (1995) interpret their results to mean that motivation, ability, and contextual cues lead some respondents to offer substantive answers rather than to choose DK, implying that a knowledge item's three response options capture an ordinal ranking. Hence, Nadeau and Niemi's perspective has much in common with my own. However, because the authors' focus is the cognitive process underlying response to knowledge items, Nadeau and Niemi discuss neither the measurement issues nor the possibility of personality-based confounds considered here.

[xiv]. Multinomial logit imposes the "independence of irrelevant alternatives," or IIA, assumption. In the current case, it is assumed that the ratio of correct answers to incorrect answers is not affected by the presence of the third choice category, DK. Intuitively, this assumption is violated. Given that

knowledge scales commonly define DK and incorrect answers as meaning precisely the same thing, it follows that if DK were not a response option, then *all* would-be DK responses should turn up as incorrect answers. To test this possibility, I reestimated the models in table 3 using grouped-data *nested* multinomial logistic regression, a technique that allows the IIA assumption to be relaxed. In these models, correct answers are the lone choice on one limb, whereas DKs and incorrect answers are the options on the second limb. These models produce an inclusive value statistic. A value near zero would indicate high correlation between DKs and incorrect answers, suggesting a) that a nested specification is appropriate, and b) in the extreme case (a score near zero), that it is proper to treat DKs and incorrect answers as indistinguishable. Conversely, a value near 1.00 would indicate that nesting is inappropriate, and, by implication, that it is improper to collapse DKs and incorrect answers into a single category. The possible superiority of the nested specification can be tested by estimating the nested model, and then reestimating it with the inclusive value constrained to equal 1.00; if the specification with the unconstrained inclusive value produces an increase in χ^2 relative to the constrained specification (critical value = 3.84 for $p < .05$), then nesting is appropriate. The values produced for this statistic are 0.13 (1992), 1.61 (1994), and 0.00 (1996). Hence, nesting is inappropriate in each year; DKs and incorrect answers are *not* interchangeable, and the two response options should not be collapsed into a single choice category.

xv. The effects in figure 2 are consistent with Converse's (1964) suggestion that the dependent variable be specified in ordinal form. Consequently, use of grouped-data ordered logistic regression may be appropriate in future analyses. The ordered model estimates only one column of coefficients because it imposes a same-slopes assumption. In the present case, we can test whether this assumption is acceptable by comparing the χ^2 statistics for the ordered and non-ordered models. If the latter exceeds the former by more than 16.92 ($p < .05$, 9 d.f.), then the ordered model should not be used. Here, the test statistics are 13.41 (1992), 6.16 (1994), and 16.53 (1996). Given that no test reaches the critical point, imposition of the same-slopes assumption in these instances would not skew results, meaning that it is acceptable (albeit only tentatively) to view the three-category knowledge measure as an ordinal scale. This constitutes strong additional evidence that DKs and incorrect answers should not be grouped together.

xvi. Neuman (1986), using data from the 1980 NES, finds that respondents with the highest levels of political sophistication have the highest tendency to claim they voted when they did not. Because validated turnout data are not available for the recent NES surveys, the results reported below must be interpreted cautiously; the actual dependent variable is whether a person claims to have voted, not whether that person truly did vote. If the effect found by Neuman is present in these latter years, it would mean that the coefficients for the knowledge variable are inflated by as much as 20 percent.

[xvii]. In Table 4, the difference in χ^2 for the two models functions as a test of whether precision is gained by distinguishing the effects of DKs and incorrect answers. The result is marginal in 1992 ($p < .07$), strong in 1994 ($p < .001$), and statistically insignificant in 1996 ($p > .10$).

[xviii]. Participants were students in an introductory American politics course at Florida State University. A self-administered, paper-and-pencil format was use. Data were gathered on the first day of class in January, 1998.

[xix]. Results consistent with those reported here were obtained when the full 21-item scale was used in place of the 15-item version. The reduced scale is used because it is more reliable than the full scale.

[xx]. The items are: “Have you usually felt pretty sure your life would work out the way you want it to, or have there been times when you haven’t been very sure about it?;” “When people disagree with you do you sometimes wonder whether you’re right, or do you nearly always feel sure of yourself even when people disagree with you?;” and “Would you say that quite often you have trouble making up your mind about important decisions, or don’t you feel you ever have much trouble making up your mind on important decisions?”

[xxi]. The risk-taking item is: “Do you think it’s always a good idea to look for new ways of doing things, or do you think in some cases it’s better to stick by what you have than to be trying new things you don’t really know about?” One bit of evidence regarding the validity of the self-confidence and risk items is that, consistent with research discussed above, male respondents on the 1958 NES exceeded female respondents in both self-confidence ($t = 4.79$) and propensity to take risks ($t = 4.39$).

[xxii]. A glimpse at the impact of personality on reliability is possible by randomly assigning DKs from the NES multiple-choice items to the available substantive response categories. In 1992, the nine

knowledge items include four short-answer measures, three two-category multiple-choice items, and two three-category multiple-choice items. The average inter-item correlation for the nine items is .325, producing an alpha of .813. When I randomly assigned DKs on the multiple-choice items to the substantive choice options, the average inter-item correlation for the nine-item scale fell to .237, producing an alpha of .737.