Appendix A

Correlational Validity

In practice, we judge correlational validity by correlating the measures we are evaluating with alternative ones, or with measures of other but related variables like education or political interest.

Closed-Ended Items

For the closed-ended case, we turn again to our TESS data. As criteria, we use political interest (measured on a five-point scale from "not at all interested" to "extremely interested"), education (categorized as less than high school, high school, some college, college, master's degree, or more advanced degree),[1] and our independent knowledge measure.

To show that our results are not merely a product of scoring decisions, we offer both the dependent and the independent knowledge indices in three different versions, all simple averages of item scores, but with the items scored differently. The first is the standard scoring, of 1 for correct and 0 for incorrect or DK. The second, a nonstochastic version of Mondak's suggestion of assigning DKs randomly to substantive response categories, scores correct answers as 1, incorrect answers as 0, and DKs as the expected value from blind guessing ($1/J$, where J is the number of response categories).[2] The third corrects for guessing by scoring correct answers as 1, DKs as 0, and incorrect answers as $-c$, where the "penalty for guessing" $c > 0$ is calibrated to bring the expected value of observed knowledge, inflated by lucky guessing, back down to that of actual knowledge. Let us call these scorings K, M^*, and G, respectively.[3]

Table A1.A suggests that closed-ended knowledge items do better to encourage than to discourage DKs (or to do neither). Under every scoring, including the Mondak-style M^*, the

DK-encouraging dependent knowledge measure correlates more highly with these criterion variables than does the DK-neutral measure, which in turn correlates more highly with them than does the DK-discouraging measure. This is also true no matter how the independent knowledge measure is scored. The average difference between the correlations with the DK-encouraging and DK-discouraging measures is .074 for *K* (the scoring that produces the highest correlations), .058 for *M**, and .056 for *G*.[4]

Recall, moreover, that DK-discouragement, for closed-ended items, is scarcely different from DK-neutrality, implying that analyses using the DK-neutral independent knowledge index as a criterion are effectively stacked (conservatively, from our point of view) in the DK-discouraging measure's favor. To the extent that the correlations are influenced by the sameness or difference of the independent knowledge measure's DK treatment, they should be highest, *ceteris paribus*, for the DK-neutral treatment, a little lower for the DK-discouraging treatment, and decidedly lower for the DK-encouraging treatment. Instead, they are highest for the DK-encouraging treatment, next-highest for the DK-neutral treatment, and lowest for the DK-discouraging treatment.[5]

Of course, higher correlations need not betoken greater validity, to the extent that less valid measures have greater extraneous covariance with given criteria. The DKs to a DK-encouraging item could for instance be more a function of political interest (as distinct from knowledge) than the DKs to a DK-discouraging one. The former's correlation with political interest could therefore be higher, even if the latter is more valid. But the independent knowledge measure, despite being stacked in favor of the DK-discouraging measure, tells the same story as the other criteria.

Open-Ended Items

This analysis can be loosely replicated for the open-ended knowledge items in the 2000 ANES. As criteria, the data again afford education, political interest, and an eight-item independent knowledge measure, plus, in this case, interviewer ratings of the respondent's intelligence and political information. The independent knowledge measure rests on a closed-ended item about the direction of any change in the federal budget deficit since 1992, two effectively closed-ended ones about party control of the House and Senate, and five open-ended ones about the home states of the four presidential and vice-presidential candidates and about Joe Lieberman's religion.[6]

The results, in Table A1.B, suggest that discouraging DKs may be a slight improvement for open-ended items. The DK-discouraging measure is more highly correlated than the DK-neutral one with every criterion but education, but the differences are generally slim, averaging only .016.[7] Again, moreover, the DK-discouraging dependent knowledge measure operates at an advantage, since five of the eight independent knowledge items (the open-ended ones) are DK-discouraging. Here this advantage puts the reality of the already small difference between DK treatments in doubt, in contrast to the closed-ended case, where it understates the much larger ones.

Table A1

Correlations with Criterion Variables

Scoring/DK Treatment	Political Interest	Education	Rated Intelligence	Rated Information	Independent Knowledge Measure *K*	*M**	*G*
A. Closed-Ended Items (TESS)							
K							
DK-encouraging	.456	.460	--	--	.492	.490	.458
DK-neutral	.373	.403	--	--	.475	.462	.441
DK-discouraging	.328	.420	--	--	.432	.435	.412
M*							
DK-encouraging	.420	.450	--	--	.465	.472	.439
DK-neutral	.341	.382	--	--	.459	.451	.430
DK-discouraging	.326	.417	--	--	.429	.433	.411
G							
DK-encouraging	.421	.453	--	--	.467	.474	.441
DK-neutral	.338	.384	--	--	.462	.453	.433
DK-discouraging	.325	.418	--	--	.431	.435	.413
B. Open-Ended Items (2000 NES)							
DK-neutral	.425	.467	.481	.561	.628	.629	.609
DK-discouraging	.486	.400	.493	.581	.687	.679	.663

NOTES

[1]"Some college" also includes associate and nursing degrees. The mean knowledge scores are nicely monotonic with these categories for all DK-treatments and knowledge scorings.

[2]"Blind guessing" here means a random draw from a uniform distribution. *M** obviously has the same expectation and on average the same sample mean as Mondak's actual recommendation of assigning DKs randomly to substantive response categories, while avoiding the latter's injection of random error. It is therefore equivalent for description and superior for correlation (Luskin and Bullock 2005).

[3]Luskin and Bullock (2005), who describe them in more detail, argue that *G* works best for description, *K* best for correlation, and *M** worst for both.

[4]Averaging the correlations with education and political interest with a pre-average of the correlations with the three scorings of the independent knowledge measure.

[5]The correlations in the table are all Pearson's *r*'s, except for those with education, which are etas. Substituting Pearson's *r*'s does not alter any conclusions: The correlations with the DK-discouraging measure are still uniformly lower than those with the DK-encouraging measure under all scorings of both the dependent and independent knowledge measures.

[6]Specifically, we gauge political interest by V001367 (scored from 1 to 4), education by V000913 (categorized as described in the text), rated intelligence by V001746 (scored from 1 to 5), and rated information by V001745 (scored from 1 to 5). The independent knowledge measure contains V001592A (whether "compared to 1992, the federal budget deficit is now smaller, larger, or about the same"), V001356 and V001357 (party control of the House and Senate), V001472 (Lieberman's religion), and V001458, V001462, V001466, and V001470 (the

home states of George Bush, Dick Cheney, Al Gore, and Lieberman). The *M** and *G* versions of the composite index differ only, perforce, in the scorings of the three closed-ended items.

[7]Again averaging the correlations with the other criteria with a pre-average of the correlations with the three scorings of the independent knowledge measure.