Let Them Guess: The Sensitivity of Political Information Effects to Variations of Question Format

Paul Weith

CEU

Abstract

Political information is generally measured by adding up the number of correct answers given by respondents to quiz-like questions related to political knowledge. The shortcomings of this approach to measuring awareness, ability or skill have been discussed in the psychometric literature for more than half a century, but only recently have they started to arouse the interest of political scientists. To mention only a few of these shortcomings (which will be discussed in this article), the questions that are used for constructing ability scales are not of equal difficulty; the format of the questions (open ended/ multiple choice/ true or false) yield different rates of correct responses by allowing the respondents to guess at unequal rates; the items may not always fit on a single dimension (which may mean that knowledge is not one-dimensional, but the more inviting conclusion is, certainly, that the data fails to conform to the empirical reality). The question that we address in this article is whether the guess related inflations of the averages of knowledge scales affect the reliability and validity of inferential statistical findings obtained with models in which knowledge is used as an independent or dependent variable in a multivariate fashion. In other words, assuming that we construct a political information variable by adding up or averaging the number of correct answers to quiz-like questions; do different question formats yield different slopes or error terms for knowledge or for its predictors in regression-based models? We start by constructing an individual-level guessing variable that accounts for the possibility and plausibility that correct answers were obtained by random guessing using data from a survey experiment carried out in Hungary in 2009. Then, we compare the results of regression models in which guessing is not accounted for with the results of nested models in which guessing is controlled for. Political knowledge is alternatively used as independent (when the dependent is the number of "don't know answers" to political attitude questions) and as dependent (for the purpose of assessing the effects of guessing on the mean of knowledge). Finally, in order to provide additional empirical support to our findings and to rule out the possibility of having results that only reflect the idiosyncrasies of the Hungarian public opinion, we use similar regression models as the ones described previously for the countries in the second module of the CSES database.

Introduction

Political information is generally measured by adding up the number of correct answers given by respondents to quiz-like questions related to political knowledge. The shortcomings of this approach to measuring awareness, ability or skill have been discussed in the psychometric literature for more than half a century, but only recently have they started to arouse the interest of political scientists. To mention only a few of these shortcomings (which will be discussed in this article), the questions that are used for constructing ability scales are not of equal difficulty; the format of the questions (open ended/ multiple choice/ true or false) yield different rates of correct responses by allowing the respondents to guess at unequal rates; the items may not always fit on a single dimension (which may mean that knowledge is not one-dimensional, but the more inviting conclusion is, certainly, that the data fails to conform to the empirical reality).

The question that we address in this article is whether the guess related inflations of the averages of knowledge scales affect the reliability and validity of inferential statistical findings obtained with models in which knowledge is used as an independent or dependent variable in a multivariate fashion. In other words, assuming that we construct a political information variable by adding up or averaging the number of correct answers to quiz-like questions; do different question formats yield different slopes or error terms for knowledge or for its predictors in regression-based models?

We start by constructing an individual-level guessing variable that accounts for the possibility and plausibility that correct answers were obtained by random guessing using data from a survey experiment carried out in Hungary in 2009. Then, we compare the results of regression models in which guessing is not accounted for with the results of nested models in which guessing is controlled for. Political knowledge is alternatively used as independent (when the dependent is the number of "don't know answers" to political attitude questions and an independently constructed knowledge scale, respectively) and as dependent (for the purpose of assessing the effects of guessing on the mean of knowledge).

While the coefficients seem virtually unaffected by question format, the possibility of guessing may still influence the distribution of error terms in our models. This would be consistent with Mondak's (2010) finding that respondents have heterogeneous propensities of guessing, which would translate into an increased heteroskedasticity for our model when guessing takes place. We test this using a heteroskedastic regression model in which we expect guessing to skew the distribution of errors if true-false and multiple choice questions are indeed less reliable than open-ended ones. Finally, in order to provide additional empirical support to our findings and to rule out the possibility of having results that only reflect the idiosyncrasies of the Hungarian public opinion,

we use similar regression models as the ones described previously for the countries in the second module of the CSES database.

The measurement of political knowledge: question format

There are certain aspects related to reliability that cannot be accounted for mathematically; the quality of measurements is always bounded by the actual data. In this regard, there are things that cannot be changed after the data was collected: the wording of questionnaire items, the format of the questions (open ended, true-false, and multiple-choice); the topics of the questions (policy-specific, ideological, campaign-specific knowledge or chronic, general), the ordering of the questionnaire items, etc. Such issues received very little attention in the field of research concerning political knowledge even though it seems apparent that this is precisely the area in which significant improvements can be made both regarding the validity and the reliability of the measurements.

This article focuses on a specific aspect of questionnaire design, while leaving further inquiries open for subsequent research. Can the reliability and/or validity of multivariate statistical findings in the study of political information effects be questioned on grounds related to question formats? Do we have reasons to encourage the use of a certain question format instead of another when constructing political information variables based on quiz-type questionnaire items?

Open ended questions tend to foster what the literature calls false negatives: a nontrivial proportion of relatively knowledgeable respondents either give incomplete answers according to the standards of the coders/interviewers (Gibson and Caldeira, 2009) or they simply shy away from giving an answer due to personal insecurity (Lupia, 2009). True-false questions and multiple choice ones, on the other hand, register false positives, meaning that a noticeable proportion of people who are not particularly knowledgeable take the opportunity to guess the answers to the questions, thus seeming more knowledgeable than they are in fact (Mondak, 2001).

These issues related to questionnaire design are particularly important due to questions of validity and reliability. It is noticeable that open ended questions systematically underestimate the level of political knowledge of respondents due to the aforementioned false negatives, whereas true-false and multiple choice questions tend to overestimate it. Furthermore, it can be argued that open-ended questions and multiple choice or true-false questions in fact measure different things: while open-ended questions test the ability of respondents to reproduce information gathered prior to the survey interview; multiple-choice and true-false questions merely capture their ability to recognize certain pieces of information that they may normally not be able to reproduce (Kubinger et al, 2010). This fact would not pose any kind of problems were the underestimations and

overestimations random – homogeneously spread throughout the sample. However, as previous findings have shown, there are statistically significant differences between those who are willing to attempt at guessing the answers to the questions and those who shy away from doing so. Generally, men, people who are less educated, those who are interested in politics and those who are more self-confident are more likely to try to attempt at guessing when they don't know the answers (Schuman and Presser, 1980; Steele and Aronson, 1995; Sturgis and Smith, 2010). Our results (see table 2) bring further corroborating evidence to these hypotheses. Moreover, the literature shows that the total amount of random guessing tends to be highest when the political interest question precedes the set of knowledge items (Sturgis and Smith, 2010). This being the case, validity and reliability cannot be assumed for measurements of political information, proper assessment and careful consideration for measurement quality should be common practice.

Extant approaches to dealing with format effects

The general purpose of the existing literature on scaling, format effects and guessing is to create reliable rankings of the abilities of test-takers based on the latent structure of their response patterns (see Rasch, 1960; Kubinger et al, 2010). The necessity for such studies can be understood best in the context of the educational psychology literature, where improving the assessment of skills and fine-tuning grading methods are particularly relevant as endpoints of the research endeavor. In this case it is rather unsurprising that little attention has been given to the performance of the estimates of personal ability (skill) in inferential statistics or to how guessing affects the conclusions we can draw from models where ability/skill is used as an independent variable. Not only does this study try to shed some light on these fairly novel issues, but it also tries to bring some important psychometric concepts and concerns closer to the political science literature.

The literature on psychometric scaling often reports large and significant differences in the rates of correct responses to test items as a direct function of question formats: for instance, Kubinger and his associates (2007; 2010) found that mathematics students score similarly on tests with *free response* format and on tests with *2 out of 5*[1] multiple choice questions but they tend to score much higher on multiple choice tests with *1 out of 6* format. The interpretation of these results is rather straightforward: in open ended questions the respondents have virtually no chance of guessing the correct answer; on *2 out of 5* questions their chance of guessing the correct answer equates to 1/10, whereas on *1 out of 6* questions this probability increases to 1/6. As the likely

---

1   There are 5 response options out of which one single combination of two is correct. The answer will be considered correct if and only if both responses are ticked.

returns from trying to guess drop, the attempts at guessing also drop, rendering the *2/5* format more similar to the open ended one in this respect.

A method of assessing political knowledge that has recently started to gain coverage in the political science literature is the Rasch model (see Selb and Lachat, 2009), which is similar in structure to the *two-parameter logistic* IRT model but differs from it significantly in method. While they both aim at estimating two latent traits underlying the performance of respondents to quizzes or tests – the ability of the respondent and the difficulty of test items – the Rasch model uses a simple logistic function that pools together both respondent and item dummies as explanatory variables predicting performance (Rasch, 1980). Rasch models do not explicitly account for the phenomenon of random guessing; nevertheless, guessing is likely to be responsible of situations in which respondents with a certain estimated level of ability provide correct answers to test items of higher difficulty (Smith, 1991). Such information is certainly not sufficient for constructing and using estimates of question-format induced guessing as controls in multivariate models.

The *three-parameter logistic* IRT model, on the other hand, utilizes explicit controls for guessing (Martin, del Pino & Del Boek, 2006); however, the standard approach is to equate the expected amount of guessing to 1/k for all test items with k response categories. Alternatives to this include accounting for the ability of the respondents or for the difficulty of test items (Martin, del Pino & Del Boek, 2006). The approach employed in this article substantively falls in the latter category, in that it takes into consideration the number of response categories and the empirical difficulty of the questionnaire items at hand. Nevertheless, the main contribution of this article is not intended for the field concerning psychometric scaling; our approach to estimating guessing has the sole purpose of creating a computationally straightforward control for format-effects that is easily applicable to the context of this particular research project.

Data, methods and results

The data that was used in this article is drawn from a phone survey experiment conducted in September 2009 on a random sample of Hungarian citizens. The experimental part of the survey consisted of randomly assigning respondents to groups differing only in the type of question they were asked. The political information items are divided into two batteries. The first battery of items consists of 6 questions, each of them having three different versions: an open-ended version, a multiple choice one and a true-false one. The last battery includes 6 variables with two versions each: one version in which the DK is encouraged and one in which it is discouraged. For each item in the first set, there was roughly one third of the sampled population that responded to each of the

three types of questions. The versions (a, b and c) were shuffled after each response, so that there was no single respondent who received only questions of type a or of type b. Similarly, the second set of questions had versions "a" and "b", with roughly 50% of the sampled population answering to one version of a certain question and roughly 50% answering to the other version.

The dataset also contains information on the respondents' level of education, their media consumption (TV news, newspapers and internet), their sex, age, income and interest in politics; all these variables were used as controls in the models. Furthermore, respondents were asked to rate their level of talkativeness, thus allowing us to use this as a control variable as well. With a sample size of 1705 and an item non-response (NA, excluding DK) rate averaging below 1% on the variables included, all the parameter estimates are likely to be statistically unbiased and representative of the target population.

If we consider that question formats should affect the reliability and/or the convergent validity of political information scales, it is reasonable to expect that knowledge variables created based on questions with different formats would associate differently with the most common correlates of political information. To this end, I constructed separate scales counting the number of correct answers to the three types of questions and checked how they correlate with traditional predictors of political knowledge: interest in politics and education. Surprisingly enough, the three types of questions perform almost equally well. In fact, the scale based on the multiple choice questions correlates with both education and interest slightly better than the other two (r=.238 with education and r=.138 with interest, whereas the other two scales correlate at .122 and .127 with interest and at .160, respectively.151 with education).

Furthermore, we intuitively expect more talkative people to be more prone to attempt at guessing when they do not know the correct answers, thus artificially increasing their political information scores. Thus, if open-ended questions are indeed conducive to false negatives (people who are fairly knowledgeable shy away from giving an answer), people who are less talkative should be less likely to provide correct answers to such questions than more talkative respondents. Hence, the association between talkativeness and the performance of respondents on open-ended questions should be informative in this respect. However, we found no relationship between talkativeness and the propensity to give a correct response to any of the open ended questions; the only logit parameter estimate that came close to traditional standards of statistical significance (p<0.08) does not even point in the expected direction.

We then created an individual level variable capturing the expected inflation of the mean of the knowledge scale that is due to the possibility and plausibility of guessing[2]. This variable was

---

2   Possibility refers to the actual chance of being able to give a correct response by guessing at random. The possibility of guessing will be equal to 1/k for multiple choice items with k response categories. Plausibility refers to how likely

constructed based on the expected results presented in Table 1: I summed up the expected mean-inflation of each multiple-choice and true-false scale (relative to the open-ended versions) that was used for the measurement of the respondents' knowledge. For instance, if a certain respondent was assigned the multiple-choice version of the first question, the true-false version of the second and the open-ended version of the third, she would accumulate 46.23-29.4=16.83 on the guessing variable for the first question, 45.9-22.9=23 for the second one and 0 for the third one, which adds up to a total score of 39.83. The score, thus, corresponds to the expected amount of inflation due to guessing on the multiple-choice and true-false questions.

This slightly deviates from the common practice in IRT literature of estimating the probability of guessing by simply dividing 1 by the number of correct answers and *distractors* in each multiple choice item (Kubinger, 2007). Instead of having a priori expectations about the inflations due to guessing, we compute the guessing variable here by taking into account the difficulty of each item as well (which is given by the percentage of correct answers in open-ended questions). Thus, we would expect no guessing inflation in multiple choice questions if their respective open-ended versions yielded 100% correct responses; but we would expect very high levels of guessing for variables in which the open-ended versions returned no correct responses.

Table1: Results obtained through three different types of knowledge measurement

Topic	Type	N	Correct (%)	Don't know (%)	Expected correct (%)	Difference (obs.-exp.)
Minister of education in Hungary	Open ended	599	29.40	67.3	-	-
	Multiple choice	577	66.40	20.1	46.23	20.17
	True/false	529	62.00	21.7	54.65	7.35
President of the EU Commission	Open ended	576	22.9	73.4	-	-
	Multiple choice	557	53.0	24.1	40.56	12.44
	True/false	572	56.8	31.1	45.9	10.9
Number of EU member states	Open ended	582	10.1	44.2	-	-
	Multiple choice	565	40.4	16.3	34.63	5.77
	True/false	558	58.6	16.8	46.65	11.95
Parliamentary support needed for Prime Minister	Open ended	563	34.3	37.1	-	-
	Multiple choice	570	24.4	9.6	53	-28.6
	True/false	572	32.0	17.5	58.4	-26.4
Budget of the EU	Open ended	592	32.4	54.1	-	-
	Multiple choice	556	36.0	14.6	50.06	-14.06
	True/false	557	27.6	16.0	58.2	-30.6

it is that a certain response is obtained through guessing; it takes into consideration the empirical difficulty of the question by taking into account the rate of correct answers given by respondents who received open-ended versions of the item at hand. For questions in which the open ended format yielded 100% correct responses, we estimate 0% guessing for multiple choice or true-false items (given that the content of the items is held constant). For questions in which the open ended format produced 0% correct responses, we will expect 100/k % correct responses for multiple choice questions that have k response categories.

MSZP on budget deficit	Open ended	578	40.3	34.9	-	-
	Multiple choice	571	43.3	25.2	51.8	-8.5
	True/false	556	40.3	22.7	58.8	-18.5

The expected percentages were calculated using the following formula: Expected = (% Incorrect in open ended + % DK in open ended – % DK in multiple choice or true-false)/ number of available options + % correct in open ended

We expect to find a significant positive association between "guessing" and knowledge; we expect to observe higher levels of political knowledge for respondents who had the opportunity to guess the answers to the knowledge items. However, the fact that the average scores of political information tend to be higher when respondents have this opportunity does not necessarily entail that the reliability or validity of different scales would be affected.

We count the correct responses given by each subject to the entire set of knowledge questions to obtain the respondents' level of political information. Using linear regression, we try to predict this score using variables often cited as predictors of political knowledge (education, political interest, gender, income, etc.) and other controls. In an alternative model, we also include the guessing variable, which is expected to be positively related to the observed level of political knowledge but also to affect some of the slopes initially found. In other words, if we construct a political information variable by adding up or averaging the number of correct answers to quiz-like questions; do different question formats yield different slopes or error terms for knowledge or for its predictors in regression-based models?[3]

A second set of models are used for this purpose. Model 3 predicts the number of "don't know" answers given by respondents to the battery of political attitude questions; the political information variable described earlier is used as independent. We then (Model 4) include the guessing variable in the specification and compare the slopes and standard errors obtained with the two models. The four models are formally expressed below:

Model 1:

Knowledge = β_0 + β_1 * interest + β_2 * education + β_3 * TV+β_4 * newspaper + β_5 * religiosity+ β_6 * family income + β_7 * age + β_8 * gender + β_9 * talkativeness + e

Model 2:

Knowledge = β_0 + β_1 * interest + β_2 * education + β_3 * TV+β_4 * newspaper + β_5 * religiosity+ β_6 * family income + β_7 * age + β_8 * gender + β_9 * talkativeness + β_{10} * guessing + β_{11} * guessing

3   The descriptive statistics for the knowledge items in our database show that true-false and multiple choice questions tend to produce higher estimates of the political knowledge of citizens; however, these differences are not fully consistent across items, which suggests that other effects (content related ones, or effects related to the different cues offered by different question formats) may be in place.

*education+ β_{12} guessing*interest + β_{13} guessing*gender + e

Model 3:

Log (E (Attitude DK)) = β_0 + β_1 * information + β_2 * interest + β_3 * education + β_4 * TV+β_5 * newspaper + β_6 * religiosity+ β_7 * family income + β_8 * age + β_9 * gender + β_{10} * talkativeness + e

Model 4:

Log (E (Attitude DK)) = β_0 + β_1 * information + β_2 * interest + β_3 * education + β_4 * TV+β_5 * newspaper + β_6 * religiosity+ β_7 * family income + β_8 * age + β_9 * gender + β_{10} * talkativeness + β_{11} * guessing+ e

The first two models were estimated using Ordinary Least Squares regression, whereas for the third and fourth we used Poisson as link function due to the distributional properties of the dependent variable. Fully consistent with our expectations, guessing was found significant in predicting political knowledge both as main effect and as part of the joint effect with political interest (see Table 2). In addition to this, the slope of political interest almost doubles in size with the inclusion of the interaction effect. This finding suggesting that part of the direct effect of interest on the knowledge that we can observe in traditional surveys takes place through the noisy impact of the random guessing of less informed respondents. The negative sign of the estimate for the interaction effect indicates that the respondents who are not very interested in politics can display significantly higher levels of knowledge compared their true score if guessing is allowed to take place. Other slopes changed too upon the inclusion of guessing in the model specification; however, these changes are small enough and substantively uninteresting.

Models 3 and 4, on the other hand, show that the inclusion of guessing as a control variable makes virtually no difference on the parameter estimates when political information is an independent variable and the count of "don't know" answers to attitude items is the dependent. None of the effects estimated in the model lost their statistical significance or changed by more than a one-digit percentage of their initial size. The intercept dropped by approximately 9% of its initial size and is no longer highly statistically significant; which is by far the most substantive change induced by the inclusion of the guessing variable in the model. The fit of model 3 is not different from that of model 4 and, most importantly, political information associates with the rate of DK answers equally well in the two models. These results seem to point to the following conclusion: the reliability and convergent validity of political information does not depend on the format of the questions used for measuring information. If there are any format effects in place, they are certainly not visible in this particular multivariate setting.

Table2: Parameter Estimates for models 1-6

	M1: Political Knowledge	M2: Political Knowledge	M3: Attitude "DK"	M4: Attitude "DK"
Political knowledge	-	-	-0.1358***	-0.1335***
Education	0.5783***	0.62406***	-0.1727***	-0.1717**
Interest in politics	1.2325***	2.2***	-0.7396***	-0.7386***
Newspaper consumption	0.1851*	0.1874*	-0.2134*	-0.2153*
TV news consumption	0.1285	0.1312	-0.2545.	-0.2508.
Guessing	-	0.0144**	-	-0.0014
Religiosity	-0.0075	-0.0075	0.0023	0.0023
Family income	-0.0007	-0.0006	0.0121***	0.0121***
Age	0.0127***	0.0132***	0.0261***	0.0262***
Gender	1. 0131***	1.2555***	-0.4140***	-0.4176***
Talkativeness	-0.9942**	-0.9170**	-0.2489	-0.2544
Guessing * education	-	-0.0007	-	-
Guessing* interest	-	-0.0145*	-	-
Guessing * gender	-	-0.0017	-	-
Intercept	-0.053	-1. 0196*	-0.9208***	-0.8374**
Model fit	R^2=0.30	R^2=0.30	AIC=2457.9	AIC=2464.5

- "." – $p<0.1$; "*" – $p<0.05$; "**" – $p<0.01$; "***" – $p=<0.001$

Further tests: CSES

It can be argued that our results are merely indicative of certain idiosyncrasies related to the Hungarian public opinion and politics. Our data, after all, was collected from a sample that is representative of the Hungarian voting age population as of November 2009. Thus, in order to provide additional empirical support to our findings we will use similar regression models as the ones described previously for the countries in the second module of the CSES database. Due to the limited number of knowledge items and the non-experimental nature of the data, the results are not expected to be as robust as the ones we obtain using the Hungarian survey experiment. Nevertheless, not only can the CSES data corroborate our results with further empirical evidence, but it can also cast some doubt on our previous findings.

The Comparative Study of Electoral Systems dataset contains a large number of variables relevant for studies in political science and is composed of representative samples from electoral democracies on all continents. Compared to other databases of similar magnitude, CSES provides considerably more diversity without making major compromises that would hinder cross-country

analyses. For the purpose of this study the second module of the database was used; the only case selection employed was lead by the occasional unavailability of the relevant variables in some of the sampled countries.

The database does not have a battery of political attitude questions; hence, the dependent variable that we use is a dichotomous one that takes the value 1 for those respondents who successfully managed to position themselves on the left-right political scale and 0 for those who failed to do so. Furthermore, there are only three political information variables for each country, and the nonexperimental nature of the data does not allow us to construct a "guessing" variable that would make this stage of the research a full replication of the previous one. Finally, some of the variables that were used at the previous stage are not available for the countries in the CSES database; namely the media consumption ones, the talkativeness variable and the interest in politics.

In some of the sampled countries only open ended political knowledge items were used (Germany, Brazil, Italy and Romania, for instance), whereas in others all three items were of true-false format (Great Britain, Hong Kong, Sweden, etc.). In most countries, however, the knowledge items were more diverse in format, thus allowing us to create relevant macro-level question format variables for subsequent use as macro-level predictors of the association between political information and the likelihood of item non-response on the left-right self-placement variable.

For the first stage of the analysis we fitted models with identical specifications to the data from each separate country in the CSES sample. With three exceptions (Russia, Romania and Mexico), the models had AIC below 1000 and the parameter estimates for political knowledge were highly significant (with the exception of Hong Kong). We then collected the slopes, standard errors and t values from the models into separate vectors (see Table 3 below). We bound the vectors together with the count of multiple-choice and true-false questions for each country and tried to predict the size of the slopes, the size of the errors and the t values using these count variables separately and together. As shown in Table 4, none of the results is significant.

Table 3. Logit parameter estimates, SEs and t-tests for knowledge

Polity	Slope for knowledge	Standard error	t-test
Albania	0.46**	0.17	2.72
Brazil	0.30**	0.10	3.07
Finland	0.63***	0.14	4.47
Germany	0.46**	0.17	2.58
Great Britain	0.81***	0.11	6.95
Hong Kong	-0.17	0.12	-1.37
Hungary	1.21***	0.20	5.82
Israel	0.97**	0.32	2.99
Italy	0.71***	0.14	5.04
Korea	0.66***	0.12	5.31
Mexico	0.30***	0.06	4.95
New Zealand	0.30***	0.10	2.99
Romania	0.39***	0.08	4.71
Russia	0.58***	0.07	7.69
Spain	0.69***	0.19	3.61
Sweden	0.69*	0.30	2.30
United States	0.44***	0.11	3.78

Table 4. OLS coefficients for question format (intercept in parentheses)

Independent / Dependent	No. of TF questions	No. of MC questions	No. of OE questions
Slopes for knowledge	-0.037 (0.58***)	0.032 (0.54***)	0.019 (0.52**)
Standard errors of the slopes	0.009 (0.144***)	0.004 (0.149***)	-0.010 (0.170***)
t-tests	-0.519 (4.350***)	0.192 (3.904***)	0.370 (3.286***)

- The first knowledge item in Hungary was coded as TF due to the high level of guessing that is allowed

We find that the slopes and errors are not sensitive to variations of format, which leads to the conclusion that it is virtually irrelevant which type of questions we use for measuring political knowledge as long as the purpose of our measurement is to use political information as an explanatory variable in multivariate models. The intercepts are highly significant in all models, suggesting that the convergent validity of the political information scales used is consistent across subsamples (polities) regardless of the heterogeneity of question formats. Furthermore, we ran analyses of variance for the categories of the TF, MC and OE variables and found no differences between the slopes of knowledge across the categories of our independent variables. The post-hoc tests revealed one significant effect[4] out of 27; which is less than what we would expect due to

---

[4] The standard error of the slope of political information is statistically significantly higher in countries where 1

chance alone.

Conclusions

The purpose of this study was to determine whether the format of the questions used for constructing political information scales influences the reliability and validity of political information variables operationalized as the count of correct answers to quiz-type questionnaire items. We tested this on experimental data from Hungary and used the CSES Module 2 database for partial replication of the results. We found no significant format effects on the slopes and errors of political information neither on the Hungarian database nor on the CSES data.

Regardless of the type of questionnaire item (i.e. regardless of the amount of guessing that the questions allow) that is used, the resulting knowledge scale is equally reliable and valid. Scales constructed based on true-false questions associate with traditional correlates of political information (political interest and education) just as well as those based on multiple-choice or open-ended questions. Furthermore, controlling for the amount of random guessing that the battery of knowledge questions allows did not affect the performance of the knowledge variable (used as independent) in any of the regression models that were tested. In this respect, we can claim that true-false, open ended and multiple choice item based scales perform similarly – in fact, almost identically – in regression models when used alternatively as independent variables. This leads us to the conclusion that most of the concerns that were mentioned earlier are not as well grounded as we would intuitively expect. These conclusions seem consistent across the countries (subsamples) in the second module of the Comparative Study of Electoral Systems.

Due to the unavailability of data for a full replication of the study presented in the first part of the present article, our tests on the CSES database are merely tentative. A full replication of the Hungarian survey experiment is necessary for a robust confirmation of our results. Moreover, while the validity and reliability of political information measurements do not seem to depend on the format of the questions used, the variations at the level of the content difficulty of questionnaire items may be more consequential. This, however, will be the topic of subsequent research.

---

multiple choice variable was used compared to the countries in which no multiple choice questions were used

References:

Boudreau, C., & Lupia, A. (2009). Political Knowledge (version 2.0). In Druckman, J. N., Green, D. P., Kuklinski, J. H., & Lupia, A. (Eds). *The Cambridge Handbook of Experimental Political Science,* Cambridge University Press (forthcoming 2011)

Gibson, J. L., & Caldeira, G. A. (2009). Knowing the Supreme Court? A Reconsideration of Public Ignorance of the High Court. *The Journal of Politics, 71*(02), 429.

Kubinger, K. D, and Gottschall, C.H. (2007). Item Difficulty of Multiple Choice Tests Dependant on Different Item Response Formats – An Experiment in Fundamental Research on Psychological Assessment. *Psychology Science,* 49(4), 361-364.

Kubinger, K et al. (2010). On Minimizing Guessing Effects on Multiple-Choice Items: Superiority of a Two Solutions and Three Distractions Item Format to a One Solution and Five Distractors Item Format. *International Journal of Selection and Assessment,* 18(1): 111-115.

Lupia, A. (1994). Shortcuts Versus Encyclopedias: Information and Voting Behavior in California Insurance Reform Elections. *The American Political Science Review, 88*(1), 63-76.

Lupia, A. (2006). How Elitism Undermines the Study of Voter Competence. *Critical Review: A Journal Of Politics And Society, 18*(1), 217.

Martin, E.S., del Pino, G., & De Boeck, P. (2006). IRT models for ability-based guessing. *Applied Psychological Measurement*, 30, 183-203.

Miller, M. K., & Orr, S. K. (2008). Experimenting With a "Third Way" in Political Knowledge Estimation. *Public Opinion Quarterly. 72(4),* 768-780.

Mondak, J. J., & Anderson, M. R. (2004). The Knowledge Gap: A Reexamination of Gender-Based Differences in Political Knowledge. *The Journal of Politics, 66*(2), 492.

Mondak, J. J. (2001). Developing Valid Knowledge Scales. *American Journal of Political Science, 45*(1), 224-238.

Prior, M., & Lupia, A. (2008). Money, Time, and Political Knowledge: Distinguishing Quick Recall and Political Learning Skills. *American Journal of Political Science, 52*(1), 169-183.

Rasch, G. (1980). Probabilistic Models for Some Intelligence and Attainment Tests; Chicago, IL, USA: University of Chicago Press.

Schuman, H., & Presser, S. (1980). Public Opinion and Public Ignorance: The Fine Line Between Attitudes and Nonattitudes. *The American Journal of Sociology, 85*(5), 1214-1225.

Selb, P., & Lachat, R. (2009). The More, the Better? Counterfactual Evidence on the Effect of Compulsory Voting on the Consistency of Party Choice. *European Journal of Political Research, 48*(5), 573-597.

Smith, R.M. (1991). IPARM: Item and Person Analysis with the Rasch Model; Chicago, IL, USA: MESA Press.

Steele, C. M., & Aronson, J. (1995). Stereotype Threat and the Intellectual Test Performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797-811.

Sturgis, P., & Smith, P. (2010). Fictitious Issues Revisited: Political Interest, Knowledge and the Generation of Nonattitudes. *Political Studies, 58*(1), 66-84.

September 2009 Computer-Assisted Telephone Survey on Political Knowledge Carried Out by the Szonda-Ipsos Polling Company for the 2009-2011 Norway Grants Project of the Hungarian Election Study.