ISSN: 1852-9003 - eISSN: 2660-700X
DOI: https://doi.org/10.14201/rlop.22687

ASSESSING POLITICAL KNOWLEDGE SCALES ACROSS COUNTRIES: EVIDENCE FROM LATIN AMERICA

La evaluación de escalas de conocimiento político entre países: evidencia de América Latina

Avaliando escalas de conhecimento político entre países: evidências da América Latina

Frederico BATISTA PEREIRA

University of North Carolina at Charlotte. USA. Email: fbatist1@uncc.edu.

Submission: 2019-10-07
Accepted: 2020-03-14
First View: 2020-05-14
Publication: 2020-05-30

Keywords:
Political Knowledge; Measurement; Cross-National Surveys; Americas-Barometer

Abstract

An informed public is viewed as essential for democratic representation. Exist- ing work suggests individual-level characteristics such as education affect political information. But contextual factors such as level of democracy and type of electoral system may also impede or facilitate the acquisition of knowledge. However, survey questions often vary across countries, making it difficult to identify the role of contextual factors. Using the AmericasBarometer surveys from Latin America and the Caribbean, this paper compares alternative scaling methods and demonstrates that the conclusions about knowledge cross-nationally depend on the measure used. The analyses compare a raw additive scale, a standardized additive scale, factor scores, item response scores with anchoring, and item response scores with bridging. The conventional additive scale suggests that very little predicts variation in knowledge across countries, while the alternatives show that factors such democracy, telecommunications, ethnolinguistic diversity, and electoral system have substantial effects on knowledge.

Palabras clave:
Conocimiento político; Medición; Encuestas transnacionales; Barómetro de las Américas

Resumen
Un público informado es percibido como esencial para la representación democrática. El trabajo existente sugiere que las características a nivel individual, como la educación, afectan a la información política. Pero factores contextuales como el nivel de democracia y el tipo de sistema electoral también pueden impedir o facilitar la adquisición de conocimiento. Sin embargo, las preguntas de la encuesta a menudo varían de un país a otro, lo que dificulta la identificación del papel de los factores contextuales. Utilizando las encuestas del Barómetro de las Américas para América Latina y el Caribe, este artículo compara métodos de escala alternativos y demuestra que las conclusiones sobre el conocimiento a nivel internacional dependen de la medida utilizada. Los análisis comparan una escala de aditivos en bruto, una escala de aditivos estandarizada, puntajes de factores, puntajes de respuesta al ítem con anclaje y puntajes de respuesta al ítem con puente. La escala aditiva convencional sugiere que muy poco se predice la variación en el conocimiento entre países, mientras que las alternativas muestran que factores como la democracia, las telecomunicaciones, la diversidad etnolingüística y el sistema electoral tienen efectos sustanciales en el conocimiento.

Palavras-chave:
Conhecimento político; medição; pesquisas transnacionais; Barômetro das Américas

Resumo
Um público informado é visto como essencial para a representação democrática. Os trabalhos existentes sugerem que as características individuais, como a educação, afetam as informações políticas. Mas fatores contextuais, como nível de democracia e o tipo de sistema eleitoral, também podem impedir ou facilitar a aquisição de conhecimento. No entanto, as perguntas da pesquisa geralmente variam de país para país, dificultando a identificação do papel dos fatores contextuais. Usando as pesquisas do AmericasBarometer da América Latina e do Caribe, este artigo compara métodos alternativos de escala e demonstra que as conclusões sobre o conhecimento internacionalmente dependem da medida usada. As análises comparam uma escala aditiva bruta, uma escala aditiva padronizada, pontuações fatoriais, pontuações de resposta ao item com ancoragem e pontuações de resposta ao item com ponte. A escala aditiva convencional sugere que muito pouco prevê a variação no conhecimento entre os países, enquanto as alternativas mostram que fatores como democracia, telecomunicações, diversidade etnolinguística e sistema eleitoral têm efeitos substanciais no conhecimento.

The extent to which citizens understand the processes and issues in politics is a key element of democratic citizenship. Politically knowledgeable citizens tend to behave in different ways from their less informed peers (Delli Carpini and Keeter 1996; Lau and Redlawsk 2001). Knowledgeable individuals are better able to navigate through the complex political processes, to properly understand and fulfill their rights and obligations as citizens, to have clear and stable preferences to be voiced and taken into account by policy-makers, and to monitor and hold political actors accountable at different levels of government.

Two important questions in the comparative study of public opinion refer to whether and why citizens from certain countries are more knowledgeable than others. However, the answers to those questions are limited by the quality of measures of political knowledge across countries. While developing comparable measures is a problem inherent to cross-national research (Przeworski and Teune 1967; Verba 1993), it is particularly hard with regard to measuring knowledge (Milner 2002). The main problem is that answers to knowledge questions usually reflect differences in the features of the questions asked across countries and over time, rather than institutional and developmental differences that affect political engagement (Elff 2009; Barabas et al. 2014). As a result, cross-country comparisons of citizens' knowledge levels are difficult and might be fraught with errors.

This paper evaluates different approaches to measuring political knowledge in cross-national surveys and shows how the choice of measure can affect the conclusions about the cross-national correlates of political knowledge. The point of reference in the analyses is the additive scale of correct answers, which does not explicitly attempt to address issues of measurement equivalence across countries. I compare this measure to its standardized version, that is, by setting its mean to 0 and variance to 1 in each country, to factor scores for each country, to an item response scale that assumes that certain item parameter is constant across countries (anchoring), and to an item response scale that assumes that certain respondents have the same underlying knowledge across countries (bridging). Using data from the 2010 round of the AmericasBarometer, I perform item analyses based on item response models to show how features of the questions rather than of the polities in which the questions are asked can affect the assessment of knowledge cross-nationally. The analyses also show that the alternative strategies tend to partially overcome those limitations, and that cross-national levels of political knowledge are associated with the country's level of democracy, average investments in telecommunications, type of electoral system, and ethnolinguistic fractionalization.

The 2010 round of the AmericasBarometer provides a unique opportunity for the purposes of the paper. The sample designs and face-to-face interviews in the survey follow very similar protocols across countries, which increases the cross-national standardization of the surveys. At the same time, due to constraints imposed by the cross-national nature of data collection, the measurement of political knowledge presents a unique combination of issues of comparability, which make it ideal for the purposes of the paper. First, like most other comparative survey projects, the survey uses open-ended instead of multiple-choice items due to difficulties in providing respondents with comparable sets of response options across countries. Open-ended questions tend to be prone to contextual biases in the recall of answers, which may vary cross-nationally (Prior and Lupia 2008; Robison 2015). Second, due to the open-ended format and difficulties in defining what constitutes a partially correct answer across different countries, the survey only distinguishes between correct and incorrect answers (Gibson and Caldeira 2009;

DeBell 2013). Third, the data does not clearly distinguish between «don't knows» and incorrect answers, which also makes the measurement of knowledge prone to the influence of arbitrary contextual features (Mondak and Anderson 2004). Last, due to the length of the multi-purpose survey questionnaire, the data includes only three knowledge items, which makes estimates from measurement models more conservative. While other comparative survey projects present some of the issues mentioned above, the AmericasBarometer offers a full combination of those features. However, while the limitations challenge the task of assessing a measurement model to correct issues of comparability, they also create the circumstances where the use of appropriate techniques to solve those issues becomes most needed.

THE PROBLEM OF CROSS-NATIONAL EQUIVALENCE

Studies that examine political knowledge across different contexts have largely relied on two types of measures from survey research. The first measure is a simple summation or average of the correct items in the battery of knowledge questions (Benz and Stutzer 2004; Gronlund and Milner 2006; Iyengar et al. 2009, 2010; Dassonneville and McAllister 2018). A second measure calculates the distance between each respondent's ideological placement of parties and the mean placement of all respondents in each country (Gordon and Segura 1997; Berggren 2001; Toka 2008; Fraile 2013). Both are not without problems. While the additive scale ignores the issue cross-national comparability, the measure of the distance between respondents' ideological placements of parties and the mean placement in their countries is not available in most comparative surveys. The latter measure also assumes that respondents perceive and use the space of the response scale in identical ways, even across different countries[1].

In what follows, I focus on the additive scale of factual knowledge items and compare it to alternative scaling strategies. There are at least two important assumptions in this approach. The first is that different items are equally capable of differentiating between more and less knowledgeable respondents with regard to their probability of answering correctly. The second assumption is that the items are equally difficult. The approach treats questions that most respondents answer correctly the same way as questions that very few can answer. Based on the assumptions, every item receives equal weight in the final additive scale. However,

1. For a discussion about interpersonal incomparability and potential problems with this assumption, see Aldrich and McKelvey (1977). For a discussion on the extent to which this measure taps the underlying construct of political knowledge, see Fortin-Rittberger (2019).

assuming that all items have equal levels of difficulty is unrealistic. Some questions inevitably touch issues and figures that are more salient than others, and giving the same score points for a correct answer in such questions results in treating equally two responses that reflect different underlying levels of knowledge. Also, assuming that all questions discriminate equally well between more and less knowledgeable respondents can be problematic. For example, an item in which two respondents with different levels of knowledge have similar probabilities of answering correctly does a poor job of discriminating knowledge levels, and therefore should not be given the same weight in the final scale. The result is that the additive scale does not take into account variation in characteristics of questions across different contexts (Mondak 1999; Elff 2009).

Taking into account the different characteristics of questions is crucial for comparative research. If questions asked within a country usually have different degrees of difficulty and discrimination, the same question asked across different countries will likely display such differences. For example, a respondent from a former British colony might be more likely to correctly answer a question about the name of the Prime Minister of the United Kingdom compared to an equally knowledgeable individual from a non-British colony. In this example, the observed measurements of the same underlying construct are not equivalent across countries, since the question has different levels of difficulty. Moreover, survey questions often have more adequate phrasings and translations in some languages relatively to others, which makes them have better discrimination of the underlying construct for respondents from some countries relative to others (Iyengar 1976; Blais and Gidengil 1993; Pérez 2009)[2].

Still, one might argue that those differences should not be discounted when estimating respondents' knowledge. If a respondent from a specific country is more likely to know the name of the Prime Minister of the U.K. than an equally knowledgeable respondent from another country, that might simply indicate that he or she is less knowledgeable about that specific piece of political information being asked about. This line of reasoning is only adequate if one is interested on measuring knowledge about that specific piece of information, but not for measuring political knowledge more generally. Specific items serve as proxies in the measurement of underlying traits, and a construct such as political knowledge can be measured by a potentially very large number of different items. Since most researchers can only use a small number of questions of political knowledge, the impact of features of those questions must be minimized in the estimation of the latent ability. If the

2. Issues of measurement equivalence often appear in the literature under different terminologies. For instance, using structural equation modelling, Davidov et al. (2018) refer to configural invariance when different groups display the same underlying factor structure. Metric and scalar invariance refer to items having similar item discrimination and difficulty across groups, respectively.

goal is to measure the broader understanding of politics, which refers to how differentiated and integrated respondents' political cognitions are (Neuman 1981; Luskin 1987), the choice of which specific questions to ask should not matter for determining which individuals are more knowledgeable than others.

THE CROSS-NATIONAL MEASUREMENT OF POLITICAL KNOWLEDGE

Since it is likely that the distributions of knowledge in each country are biased because of features of the items, one approach to eliminating bias would be eliminate the cross-national variation itself. The first approach investigated here is the simple standardization of the additive scale in each country. This procedure involves transforming the average number of correct responses in each country to zero, and setting the variance of each distribution to equal 1. Even though this approach eliminates the cross-national variation in knowledge that is the object of interest and can be used in within-country analyses, it is possible that the multivariate analyses of the variation can still reveal substantial difference in the shapes of the distribution a across countries. Moreover, once the analyses control for individual-level variables that may vary across countries, this approach could potentially uncover cross-national that is not associated with individual-level variables. While the main advantage of this scale is that it is computationally simple to calculate, it does not address directly the main problems in the additive scale.

The second strategy to overcome problems in the additive scale is to use factor analysis to address differences across items. Factor scores are estimated by weighing the different items based on how strongly they correlate with (or load on) the common variance (Costello and Osborne 2005)[3]. The factor loadings indicate the extent to which each item discriminates respondents with respect to the underlying trait, which means that this approach does not consider how items can have different levels of difficulty. Moreover, in order to identify factor models, the most common constraint is to center and standardize the estimated scores. Hence, factors scores share some properties of a standardized scale (mean of 0 and variance of 1 in every country), but calculate the individual scores by weighing each item differently in each country. Like the additive and standardized scales, this approach also removes the cross-national variation by standardizing the scale in each country, while also not taking into account differences in difficulty across items. Finally, constraining the predicted scores to be standardized in the estimation of the factor loadings does not solve the issue of measurement comparability, which

3. The paper focuses on the computationally simpler exploratory factor models.

denotes that factor scores are not designed to be in a common space of measurement across countries.

Item Response Theory (IRT) offers an alternative framework for the measurement of test items (Johnson and Albert 1999; Embretson and Reise 2000). An item response model specifies that the probability of a respondent giving a correct answer to a question depends both on his/her ability to answer the item and on properties of the items being asked about. In this framework, variations in features of items become object of analysis that are taken into account when estimating individuals' abilities. In the case of political knowledge questions, the probability of observing a correct response is a function of the item characteristics (difficulty and discrimination) and the individual latent trait (political knowledge)[4].

Item difficulty indicates how knowledgeable a respondent must be in order to give a correct answer to the question. Questions with higher values in this parameter are considered harder, since respondents must reach a certain threshold of knowledge in order to answer correctly. Item discrimination corresponds to the extent to which a question discriminates between more and less knowledgeable respondents. Both item difficulty and discrimination parameters are given in the same metric as the estimated latent ability. Moreover, the estimation of levels of difficulty and discrimination is simultaneous and conditional on the latent trait. In this sense, the approach estimates the parameters for different items while holding constant the underlying levels of knowledge of respondents. As a consequence, the approach allows questions' characteristics to be estimated and taken into account in the calculation of respondents' political knowledge scores. Respondents who fail to give a correct answer to a difficult question do not receive the same penalty as respondents who fail to answer an easy one. Similarly, respondents that do answer the harder question correctly receive a higher reward than respondents who answer the easier one. Also, questions that are better at discriminating more from less knowledgeable respondents receive more weight in the estimation of the knowledge scores than questions with low discrimination.

Although item response models offer a more flexible approach in comparison to the additive scale for dealing with the issue of item equivalence in cross-national survey research, they also rely on measurement assumptions that impose identification constraints to their scales. Two main types of constraints are generally used in psychometrics. The first is setting the mean and the variance of the estimated latent ability (either for each country or for all countries) to specific values, usually zero and one, respectively[5]. The second common constraint

4. This paper focuses on the specific case of the two-parameter logistic model. Examples of alternative models are the Rasch model, in which only item difficulty is estimated, and the three-parameter model, that also estimates a «guessing» parameter (lower asymptote) for each item.
5. This is analogous to identification constraints in factor analysis models.

is to set the discrimination parameter for one item to be necessarily positive (or negative, depending on the goals of the researcher), which solves the problem of rotational indeterminacy (when both positive and negative values solve the system of equations). While such restrictions are sufficient for the estimation of an identified item response model within a country at a single point of time, such as for the factor scores discussed above, they are still problematic in comparative research. The main issue is related to the attempt to solve the problem of scale indeterminacy (defining the actual range of values the estimates will take on). Setting the underlying knowledge levels to have a mean of zero and a variance of one results in setting to constant values quantities whose variation might be of primary interest in comparative research.

The first strategy to establish equivalence in IRT models is to use of anchoring items. In this case, the researcher uses one or more items that are assumed to be comparable across respondents from different groups in order to estimate the remaining item parameters and the ability scores for all individuals across the different groups (Bafumi and Herron 2010; Tausanovitch and Warshaw 2013). Measurement equivalence is established by the extent to which the assumption about the measurement invariance of the anchoring item holds. For instance, if one item displays the same difficulty or discrimination across groups, fixing that parameter identifies the item response model without sacrificing measurement equivalence. This solution is feasible when researchers have a large poll of items. However, cross-national surveys rarely include batteries of questions that are large enough to allow that.

The last approach for solving for the trade-off between identifying the measurement model and obtaining comparable country knowledge distributions that actually differ from each other lies on fixing (or imposing linearly independent restrictions on) two other points of the scale (rather than the mean and standar deviation). In other words, one must identify respondents in each group (country) that can be assumed to have the same level of the underlying ability. Scholars that study ideal points of legislators from roll call data face a similar trade-off, because since legislators from different legislatures do not cast votes on the same roll calls, using anchoring items (roll calls) is not an available solution. In order to obtain comparable ideal point estimates across different roll calls, Clinton et al. (2004) constrain the ideal points of two legislators at arbitrary positions across different votes. The two legislators are assumed to be at opposing extremes in the scale and expressing the same underlying level of ideology across different votes. Therefore, solving the comparability problem in the estimation of political knowledge scores across countries involves finding those observations (respondents) from different countries that can be used as «bridges» in the estimation of the item response models. Because those individuals are assumed to have the same underlying level of political knowledge, constraining their ability levels in the estimation provides

comparability and at the same time identifies the item response model in each country[6] [7].

The scaling strategies described above (from the additive scale to the IRT scale with «bridging») present increasing levels of computational complexity. This is so because, although the scales rest on identification constraints that make assumptions about measurement equivalence, the more computationally complex ones make fewer assumptions than the others. With fewer assumptions, those models estimate more parameters across countries. Table 1 shows the equivalence assumptions about scores and item parameters that are assumed to be invariant across countries for each scaling method discussed here:

Table 1. Equivalence Assumptions of Alternative Measurement Approaches

Scale	Invariant Scores	Invariant Items
Additive Scale	Yes: all	Yes: all
Standardized Scale	Yes: mean and SD	Yes: all
Factor Scores	Yes: mean and SD	Yes: difficulty
IRT Anchoring	No	Yes: at least one
IRT Bridging	Yes: bridges	No

As Table 1 shows, the computationally simpler approaches tend to rely on more assumptions about the cross-national comparability of scores and item parameters than the more complex ones. The latter attempt to achieve equivalence by proposing new assumptions (anchoring and bridging) that would arguably be more realistic than the conventional ones[8]. In the next section I evaluate the conventional assumptions about item invariance and standardizing distributions and examine the extent to which the IRT solutions present improvements relative to the simpler scales.

6. The approach developed here is similar to the one proposed by Lo et al. (2014) for ideology.
7. The Bayesian approach offers computational advantages in this respect. By using Monte Carlo (MCMC) simulations and by taking into account a priori distributions of parameters, the approach facilitates the estimation that would be complicated under the classical approach (Clinton et al. 2004; Jackman 2009).
8. Both IRT approaches based on Bayesian models can have some degree of standardization in the predicted scores due to the use normally distributed (uninformative) priors.

MEASURING POLITICAL KNOWLEDGE COMPARATIVELY

This section assesses the measurement approaches discussed above with knowledge questions from the 2010 round of the AmericasBarometer surveys, by the Latin American Public Opinion Project (LAPOP). The survey interviewed 43,990 pondents in nationally representative samples of 26 countries across the Americas between January and August of 2010. The questionnaire included 3 questions of political knowledge. The questions were not asked in the United States and Canada, so these countries are not included in the analyses of the paper. The remaining 24 countries have a total of 40,990 respondents. The AmericasBarometer's survey instrument and methodology are highly standardized across the 24 participant countries. The sample designs follow similar protocols across countries, and all interviews are face-to-face. In all 24 countries the political knowledge questions have the approximate same placement in the questionnaire, and all interviewers in the project are instructed to follow the same protocols. Since one of the main goals of this paper is to examine the extent to which features of the knowledge items (rather than features of the data collection process) affect their performances across different countries, the standardization of the AmericasBarometer surveys offers an ideal dataset for assessing the influence of those factors in the analyses.

The first knowledge question asked respondents to say the number of provinces (states/departments) in their country (54% of correct responses overall). The second asked about the length of the presidential/prime ministerial term in the country (81% of respondents answered correctly). The last question asked respondents to say the name of the President of the United States (74% correct). Overall, about 8% of respondents answered all questions incorrectly, 19% answered at least one correctly, 33% answered two, and approximately 41% answered all questions correctly[9].

«Don't know» (DK) answers were coded as incorrect. The extent to which DK answers indicate simple lack of knowledge (Luskin and Bullock 2011) or other attributes (Mondak and Anderson 2004) is a matter of debate among scholars. This paper focuses on the conventional measurements of knowledge that code DK's as incorrect. Additionally, this choice is conceptually related to the idea of «knowledge in use» (Neuman 1981, 1240), that is, pieces of information that individuals actually use when they think about politics, rather than the broader idea of knowledge as information individuals have but might not be willing to use or display[10].

9. The number of items in measurement models tends to increase the reliability of estimates. Therefore, using three rather than a larger number of items yields more conservative (but still valid) results.
10. There are two main reasons why the analysis does not estimate the «guessing» parameter for each item across countries. First, according to the literature, guessing is more common with close-ended

The knowledge items in the AmericasBarometer allow the construction of the five alternative scales discussed in the previous section. The additive scale is obtained by simply adding the number of correct responses given by respondents across all countries. The standardized scale consists of a linear transformation of the additive scale performed separately by country (setting the mean correct scores in each country to 0 and its standard deviation to 1). The factor scores are estimated based on maximum likelihood factor analysis models ran separately in each country. The IRT scale using anchoring is estimated by setting the difficulty parameter for the item about the name of the U.S. President (the only item with the same answer in all countries) to be constant across all 24 countries.

The IRT scale with «bridging» requires selecting «bridging» cases, that is, assuming that some respondents from all countries have the same underlying level of knowledge. An examination of the individual-level correlates of the knowledge items in each country of the AmericasBarometer 2010 shows that respondents' level of education is the only variable to have positive and statistically significant effects on all three items across the 24 countries. Therefore, the criteria used here for selecting the individuals at the bottom of the distribution of the knowledge scores is having zero years of education and having answered none of the knowledge questions correctly. In the other extreme, the criteria for being at the top of the scale is having more than 14 years of education, while also having answered all knowledge questions correctly[11]. The number of individuals meeting the criteria varies by country, and all of them in each country were used as «bridging» cases (see appendix table A1). The lower «bridging» cases are set to have a value of -2 in the underlying scale, while the higher «bridging» cases are set to a value of 2[12].
To what extent do the three knowledge items in the AmericasBarometer have the same characteristics across countries? As discussed above, both the raw and the standardized additive scales assume that the items have the same levels of difficulty and discrimination across contexts. The scale based on factor scores assumes that item difficulty is the same across countries, while the IRT scale based on anchoring assumes that either difficulty or discrimination does not vary across countries for one of the items. Therefore, the only approach that allows all item characteristics to vary across all countries is the IRT scale with «bridging». Figures 1, 2,

questions and when DKs are stimulated, which is not the case here. Second, a 3-parameter model with fewer identification constraints would make estimates less reliable.

11. Even though this identification strategy follows Clinton et al. (2004), it relies on weaker assumptions, since Clinton et al. (2004) fix the underlying ideology of the same legislators over time, while I fix the knowledge of different respondents across countries. While the assumption is unrealistic, the goal is to compare it to scales that rely on different (and potentially less realistic) assumptions.

12. I implemented this procedure and estimated the IRT models using the R package «pscl» (R version 3.5.3). I used Stata version 13.1 for data preparation and final analyses.

and 3 below show the estimates of item difficulty and discrimination for the three questions based on that item response model:

Figure 1. Difficulty and Discrimination Parameters for «Number of Provinces/States»

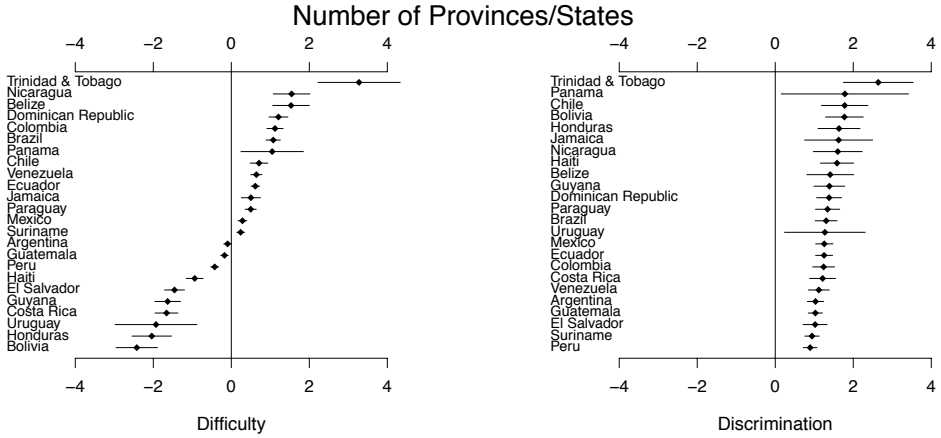


Figure 2. Difficulty and Discrimination Parameters for «Length of Presidential/ Prime Ministerial Term»

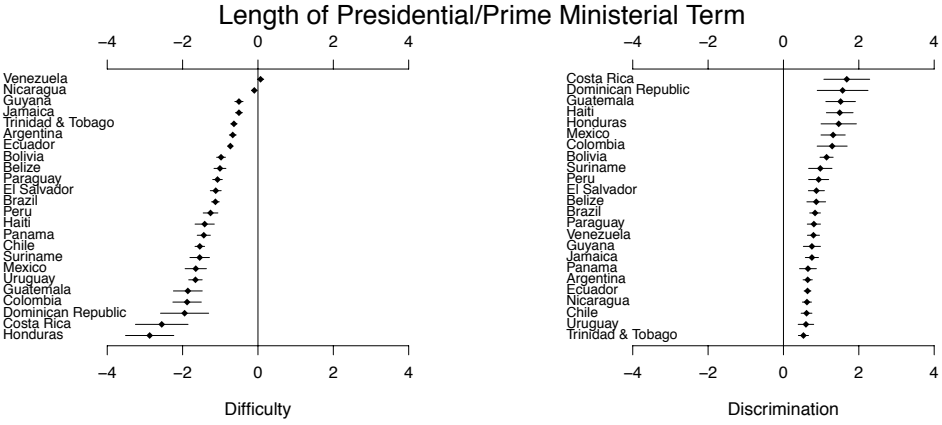
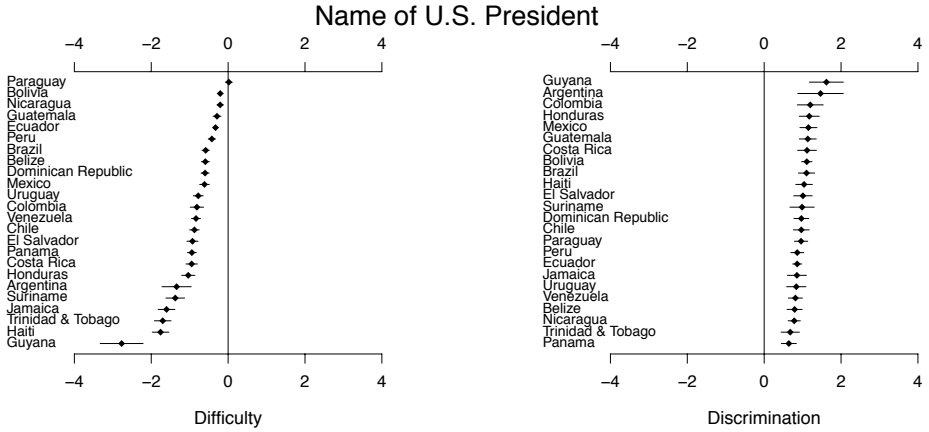


Figure 3. Difficulty and Discrimination Parameters for «Name of U.S. President»



Source: The AmericasBarometer (2010) by the Latin American Public Opinion Project (LAPOP), www.LapopSurveys.org.

The left side of the figures displays substantive variation in the estimated levels of difficulty. In Figure 1, the plot shows that the question on the number of states has large differences in item difficulty across countries. For that item, 13 estimates are statistically different from the average in difficulty (5 with the conservative Holm-Bonferroni correction). The question on the length of the presidential/prime ministerial term (Figures 2) displays less variation in item difficulty across countries than the first one, but it still shows some substantial differences. A total of 10 estimates for the item are statistically different from the average (5 with the Holm-Bonferroni correction). The question about the U.S. President shows lower cross-national variation in item difficulty. For item on the U.S. President, 10 estimates are statistically different from the average (3 with the Holm-Bonferroni correction). Hence, the results suggest that an anchoring approach that assumes that the item difficulty of item on the U.S. President could potentially produce better measurement than other approaches. Finally, the plots on the right show that there is moderate variation in discrimination parameters across countries. Overall, the items tap well the underlying construct, since the estimates for all countries are different from zero. Nevertheless, some of the questions in specific countries deviate considerably from the average item discrimination, and these discrepancies must be taken into account when estimating individual scores.

Why are items more difficult in some countries than in others? Table 2 shows that the differences in item difficulty across countries are correlated with features of the questions by regressing the estimates displayed above against some of those

features[13]. Since the estimated item difficulties for each country are the dependent variables in the analyses, the models take into account the uncertainty of those estimates. In order to do so, I use the weighting matrix proposed by Borjas and

Table 2. Two-Stage Estimates for Cross-National Correlates of Item Difficulty

Independent Variable	Number of Provinces	Length of Term	U.S. President
Number of provinces (answer)	0.05*		
	(0.02)		
Age of state/province divisions	-0.01		
	(0.01)		
Answer multiple of 10 (yes=1)	-0.82		
	(0.56)		
Reelection (yes=1)		0.70*	
		(0.24)	
Term longer than 4 years (yes=1)		0.62*	
		(0.24)	
Months since last election		0.01*	
		(0.01)	
International tourism, 2009			-0.00
			(0.00)
Merchandise imports, 2009			-0.02*
			(0.01)
Constant	-0.37	-2.04*	0.25
	(0.56)	(0.25)	(0.47)
N	24	24	24
Adjusted R^2	0.15	0.40	0.14
Residual Standard Error	1.20	0.46	0.52

*p<0.05. Standard Errors in Parentheses.

Data for dependent variables: The AmericasBarometer (2010) by the Latin American Public Opinion Project (LAPOP), www.LapopSurveys.org.

Data for independent variables: Various sources (described in the text).

13. Information about the number of provinces and the duration of the term are available in the survey. Data on territorial changes comes from Statoids (http://www.statoids.com). Election dates come from the Political Database of the Americas (http://pdba.georgetown.edu/). Data on international tourism and imports comes from the World Bank (http://databank.worldbank.org/data/home.aspx).

Sueyoshi (1994), that weights the second-level regression estimates based on the error components from the first-level IRT models and the second-stage country-level regression models[14]. Moreover, the dependent variables are the estimated levels of difficulty for each item based on the item response model described above. Difficulty parameters are estimated conditionally on levels of knowledge, and therefore the results show how features of items and countries affect the former, not the latter.

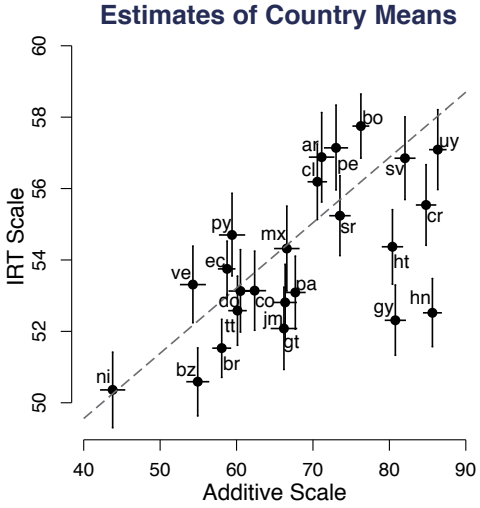
The model in the first column shows that it is more difficult for respondents in countries with more provinces/states (ranging from 7 to 60) to answer the question correctly. In countries where the correct answer corresponds to a multiple of ten, it is also easier to answer the question, but the effect is not statistically significant (only 2 out of 24 countries have round numbers as answers). The age of territorial divisions does not have a substantive or significant effect on difficulty. The question on the length of the presidential/prime ministerial term tends to be more difficult in countries with reelection and with terms longer than 4 years. It is also easier for respondents to get the correct answer in countries where the survey was conducted closer to a national election. Finally, for the question about the U.S. President, results show that it is easier in countries with larger proportions of foreign imports[15].The results show that the selection of questions arbitrarily benefits respondents from some countries, since performance in the items is largely affected by factors related to the content of the items rather than to social and institutional factors that facilitate learning about politics. Also, the fact that these estimates are systematically correlated with features of the items corroborates the validity of the estimates obtained from the IRT approach that relies on «bridging».

We also observe changes in the cross-national levels of political knowledge when we compare the additive and the IRT scales. Since the items tend to be easier in some countries, one should observe changes when using the item response models that minimize the influence of those factors. While the individual-level scores from both scales are highly correlated (0.87, $p<0.01$), the correlation between the country means is lower (0.61, $p<0.01$). These differences become clearer by considering the changes in the country rankings between the two measures. Figure 4 compares the country means in each of the knowledge scales (rescaled from 0-100)[16]: Only 3 out of 24 countries remain in the same positions across the two

14. This approach is also discussed in Lewis and Linzer (2005) and other papers.
15. Even though one could argue that trade has a positive effect on knowledge, the estimates show the effects of trade on item difficulty. The measure of foreign imports is not related to item difficulty for the items on provinces and duration of term, as well as to different scales of knowledge (see appendix table A2). This indicates that, in this case, trade is a proxy for proximity to the U.S. that affects item difficulty rather than knowledge.
16. While means range from 44 to 86 points in the additive scale, they go from 50 to 58 in the IRT scale. This shrinkage is due to the functional form used in the IRT model (similar to a factor analysis

variables: Suriname (8th), Trinidad and Tobago (18th), and Nicaragua (24th). Most countries gain or lose only a few positions, but others change dramatically their place in the ranking. Paraguay (from 19st to 9th) and Venezuela (from 23th to 13th) are the two countries that gain most positions in the IRT scale. Honduras (from 2nd to 19th) and Guyana (from 5th to 20th) make the opposite movement.

Figure 4. Country Means: Additive Scale versus IRT scale



Data: AmericasBarometer Surveys, 2010.

CORRELATES OF POLITICAL KNOWLEDGE ACROSS COUNTRIES

Comparative research provides different explanations for why citizens from some countries are more politically engaged than citizens from others. Economic development is hypothesized to increase levels of industrialization and urbanization, which improve communications, technology, and mobility. These factors facilitate the flow of information available to citizens (Lipset 1959; Castells 1996). Democracy also increases citizens' access to information by promoting freedom of association to form groups that collect, exchange, and spread information. Democratic media and political parties also have more freedom to produce and distribute politically relevant information. Additionally, direct participation in policy decisions tends to increase citizens' political knowledge and sense of efficacy (Bowler

model). Hence the numerical values in the two scales are not directly comparable.

and Donovan 2002; Benz and Stutzer 2004). Studies also document that more ethnically diverse societies tend to display lower levels of trust and civic engagement (Knack and Keefer 1997; Alesina and Ferrara 2000; Anderson and Paskeviciute 2006), which in turn diminish the degree of exchange and spread of political information. Moreover, ethnic diversity is often associated with language differences that might undermine the efficient flow of information across groups and the ability of governments and other political groups to effectively communicate with citizens. Regarding electoral systems, some scholars argue that the level of proportionality increases citizens' knowledge by fostering party differentiation and, as a consequence, more diversity of information conveyed to the public by competing sources (Gordon and Segura 1997; Berggren 2001; Gronlund and Milner 2006). The perspective offered by political psychologists suggests that the relationship could be the opposite, as a consequence of «cognitive overload». This perspective suggests that features such as large number of parties and sources of information, and more complex electoral rules can decrease knowledge levels. (Kuklinski et al. 2001; Sniderman and Levendusky 2007)[17].

A cross-nationally comparable and valid measure of political knowledge should be able to capture the variations hypothesized by the scholarship on the topic[18]. Even though the testing the associations between the different scales and theoretically related factors does not constitute an ideal test of measurement quality, it provides initial insights about the extent of their construct validity (Zeller and Carmines 1980). Table 3 shows the results from multilevel models with the effect of the theoretical correlates of political knowledge at the country-level on the alternative scales discussed in the previous section[19]. I include the GDP *per capita* in 2009 and the average investment in telecommunications from 2000 to 2009 (in ten million of dollars) as measures of economic development[20]. The inverted Freedom House Index of political rights and civil liberties in 2009 is used as a measure of level of democracy[21]. I use the Gallagher Index of proportionality (Gallagher 1991; Carey and Hix 2011) as measure of dispersion of power promoted by the country's institutions[22]. Finally, I use an average of the indexes of ethnic, linguistic,

17. Rainey (2015) also casts doubt on the positive effects of proportionality in a similar vein.
18. Although including features of the items as controls in a cross-national model partially solves the problems in the additive scale (see appendix table A4), that model does not yield the same results as the item response scale. An alternative strategy is to use the respondent-item as unit of analysis while including relevant item-related control variables (Fortin-Rittberger 2016; Batista Pereira 2019). While doing so increases degrees of freedom and allows the inclusion of controls, researchers often cannot know in advance which features to control for in order to fully eliminate bias.
19. The models take into account weights that equalize the sample sizes of all countries.
20. Both measures come from the World Bank (http://databank.worldbank.org/data/home.aspx).
21. Data from the Freedom House website (https://freedomhouse.org/).
22. Values for proportionality are missing for Panama and Haiti, which receive the mean value. The results do not change when those countries are excluded.

and religious fractionalization with data provided by Alesina et al. (2003). The dependent variables are all recoded to range between 0 and 100. The models also include individual-level controls associated with knowledge.

Table 3. Multilevel Linear Models of Correlates of Political Knowledge

Independent Variable	Additive Scale	Standardized Scale	Factor Scale	IRT Anchoring	IRT Bridging
Country-level:					
Democracy	2.54*	.06	.06	.08	.57*
	(1.07)	(.11)	(.11)	(.13)	(.17)
Telecommunications	-.01	.01*	.01*	.01*	.01*
	(.01)	(.00)	(.00)	(.00)	(.00)
Proportionality	.22	.25*	-.25*	-.30*	-.30*
	(.40)	(.06)	(.06)	(.07)	(.06)
Fractionalization	-7.14	-1.20	-1.12	-1.45	-3.85*
	(5.91)	(.83)	(.82)	(1.00)	(.91)
GDP *per capita*	-1.34*	-.51*	-.50*	-.60*	-.70*
	(.61)	(.07)	(.08)	(.09)	(.10)
Individual-level:					
Wealth	2.86*	1.73*	1.72*	2.04*	1.92*
	(.24)	(.13)	(.13)	(.15)	(.13)
Employed (yes=1)	3.65*	2.16*	2.12*	2.53*	2.38*
	(.70)	(.39)	(.40)	(.45)	(.40)
Education	2.12*	1.26*	1.27*	1.51*	2.34*
	(.13)	(.06)	(.06)	(.08)	(.14)
Urban (yes=1)	3.86*	2.44*	2.77*	1.76*	
	(.77)	(.47)	(.44)	(.52)	(.52)
Age	.06*	.04*	.03*	.04	.11*
	(.03)	(.02)	(.02)	(.02)	(.02)
Male (yes=1)	4.84*	2.88*	2.98*	3.41*	2.96*
	(.69)	(.41)	(.40)	(.47)	(.40)
Media exposure	3.90*	2.33*	2.40*	2.69*	2.17*

Independent Variable	Additive Scale	Standardized Scale	Factor Scale	IRT Anchoring	IRT Bridging
	(.42)	(.25)	(.28)	(.30)	(.25)
Political interest	1.62*	1.00*	.93*	1.18*	1.23*
	(.32)	(.18)	(.19)	(.22)	(.20)
Internal efficacy	.51*	.33*	.31*	.40*	.37*
	(.11)	(.07)	(.08)	(.08)	(.08)
Constant	15.43	43.78*	42.53*	31.35*	17.73*
	(14.80)	(2.05)	(2.11)	(2.48)	(2.48)
Respondents	38,384	38,384	38,384	38,384	38,384
Countries	24	24	24	24	24
Variance Explained (Constant)	.17	.35	.34	.34	.37

*p<.05. Standard errors in parentheses.

Source: The AmericasBarometer (2010) by the Latin American Public Opinion Project (LAPOP), www.LapopSurveys.org.

Data for country-level variables: Various sources (described in the text).

The coefficient for level of democracy on all five scales is positive, but only statistically different from 0 for the additive scale and the IRT scale with «bridging». Investments in telecommunications have a positive and similar association with all scales but the additive scale. Contrary to expectations, the coefficient for the additive scale is negative and statistically significant at 0.10. The patterns are mixed for features of the electoral system. Electoral system proportionality has a negative coefficient for factor scores and the two IRT scales. The coefficients are positive for the additive and the standardized scale, and only statistically different from 0 for the latter. While the results are mixed, additional evidence favors the negative effect of electoral system proportionality found in Table 3 using the IRT scales. First, a similar model having the measure of internal political efficacy as dependent variable finds that proportionality has a negative effect, controlling for the same factors used in Table 3[23]. This means that in more proportional systems citizens-believe they know less about politics than their counterparts from majoritarian systems. Second, Zechmeister and Corral (2013) find that increasing the effective

23. The item of internal efficacy asks respondents whether respondents agree with the statement «you feel that you understand the most important political issues of this country».

number of parties (a feature associated with proportionality) in Latin America decreases citizens' propensity to locate themselves in the left-right scale, which also suggests that higher complexity at the system level affects citizens' ability to make sense of politics. Fractionalization has the expected negative coefficient on levels of political knowledge for all scales, but the effect is more precise and statistically different from 0 for the IRT scale with «bridging.» There is a negative coefficient of economic development (GDP *per capita*) for all knowledge scales. While this effect contradicts the hypothesis that level of development increases knowledge across countries, it cannot be attributed to a particular measurement approach.

All in all, the results on Table 3 converge on the conclusion that a lower portion of the cross-national variance in the additive scale is explained by theoretically relevant factors. While the IRT scale with «bridging» tens to be slightly better than the others in recovering more precise coefficients for the theoretically relevant country attributes, the more striking pattern on Table 3 comes from the comparison between the additive scale and the other four strategies[24]. Those factors explain 16% of the country-level variable for the additive scale, while explaining between 34% and 37% for the four other scales[25]. Additionally, as suggested by the findings of the previous section, the summation of correct responses in items that have different characteristics from each other and across countries might simply transfer systematic measurement error to the resulting additive scale. In fact, some of those factors still have strong associations with the additive scale, while having much smaller and less precise associations with the other scales (see appendix table A3). The country-level variance explained by those factors with the additive scale is 17%, while it is of approximately 9% for the IRT scale with «bridging», for example.

CONCLUSION

This paper assesses the comparability of political knowledge questions in cross- national surveys and evaluates different scaling methods on how they address the issue. Item analyses show that features of items, rather than of polities, affect the cross-national variation in respondents' performance in those questions.

24. The proportion of «bridging» cases in each country does not drive the results. The same models separated between the «bridging» and the remaining cases show that the results are driven by the non-bridging cases. The fact that the «bridging» cases are not affected by the cross-national correlates evidences that assuming that they are similar across countries is adequate and fosters the equivalence in the final scale. See appendix table A2.

25. Explained country-level variation here corresponds to the country-level residuals from the final model compared to models including only the individual-level variables.

Hence, the raw additive scale of political knowledge often used in cross-national survey research can present issues of measurement equivalence. Moreover, I compare the additive scale to four alternative scaling methods: the additive scale standardized by country, predicted scores from country-specific factor analysis models, an item response scale that uses anchoring items, and an item response scale that uses «bridging» observations to establish comparability. The results show that the additive scale (both raw and standardized) is more strongly associated with arbitrary features of questions than with theoretically relevant cross-national variables, such as telecommunications, electoral system proportionality, and ethnic fractionalization. The alternative scales tend to display similar performance, with the item response scale based on «bridging» showing slightly better results.

The alternative scales (factor and item response scores) are better able than the conventional additive scale to capture substantive cross-national variation in polit- ical knowledge, and improve the extent to which scholars can assess the causes and consequences of political knowledge at the polity level. With respect to the effect of democracy and press freedom on the availability of information that citizens can learn, both the additive and the item response scale show a positive effect, in line with previous studies using the additive scale and the party-placement measure (Fraile 2013; Schoonvelde 2013). While the literature on political knowledge does not examine the effects of ethnic, religious, and linguistic cleavages, the model with the item response scale shows that those factors tend to reduce cross-national levels of knowledge, in accordance with the literature about trust and civic engagement (Knack and Keefer 1997; Alesina and Ferrara 2000; Anderson and Paskeviciute 2006). The results are mixed with respect to the effect of economic development. All scales retrieve a negative effect for the GDP *per capita*, which contradicts the expectations and findings from previous literature (Mondak and Canache 2004)[26].

On the other hand, investments on telecommunications, a more specific indicator of the impact of development on informational flow, shows the expected positive relationship for all scales but the additive. Moreover, and contrary to some views about the effect of electoral institutions on civic engagement, electoral system proportionality has a negative effect on political knowledge. The studies using the party-placement measure show that proportionality increases knowledge (Gordon and Segura 1997; Berggren 2001; Gronlund and Milner 2006), but Fortin-Rittberger (2019) shows that the measure is not strongly associated with factual knowledge question, which casts doubt on the extent to which it taps the underlying construct. Using an additive scale, Fraile (2013) finds that the effective

26. The difference relative to the study by Mondak and Canache (2004) could be due to the fact that they use a measure of knowledge of science and the environment, and not politics. Moreover, their sample includes only developed countries.

number of parties has a positive effect for countries with fewer parties, but the effect becomes negative for countries with more than 5 effective parties. Since Latin American party systems tend to have large numbers of parties, the results in Fraile (2013) are consistent with the idea that features associated with larger complexity tend to decrease knowledge in the region.

The analysis of bias in the items of political knowledge used for cross-national comparisons has implications that go beyond the understanding about what makes countries more knowledgeable. Cross-nationally biased scales of political knowledge could affect conclusions from cross-national models that have political knowledge as individual-level independent variable. Given that measurement error in the independent variable tends to attenuate the regression coefficient (King et al. 1994, 163), estimates from multi-level models including both knowledge and cross-national factors as relevant independent variables could be biased if the mismeasured country-averages of political knowledge are correlated with those cross-national predictors. Future research is required about the more general problem of how issues of measurement equivalence in individual-level variables could bias relevant group-level estimates in multilevel models. On the other hand, the issues discussed here do not affect directly the use of standard scales of political knowledge within countries, since the different scales are strongly correlated at the individual-level. Moreover, the computational complexity of item response models for measurement across groups makes it costly for single-country studies.

Finally, the results point to possible recommendations for the design of knowledge questions in cross-national surveys. First, given that arbitrary features of items will inevitably produce bias in favor of respondents from some countries, using longer and diverse batteries may contribute to minimizing those problems. Second, as the analyses of the paper show, items with the same answer across countries (such as the name of the President of the U.S.) likely have less variation in difficulty and discrimination across countries. Finally, the use of standardized questionnaire protocols and questions formats that facilitate cross-national comparisons (such as multiple choice questions) may help to minimize the extent of bias in assessments of cross-national levels of political knowledge.

REFERENCES

Aldrich, J. H. and McKelvey, R. D. (1977). A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections. *The American Political Science Review*, 71(1):111-130.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., and Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, 8:155-194.

Alesina, A. and Ferrara, E. L. (2000). Participation in Heterogeneous Communities. *The Quarterly Journal of Economics*, (w7155):847-904.

Anderson, C. J. and Paskeviciute, A. (2006). How Ethnic and Linguistic Heterogene- ity Influence the Prospects for Civil Society: A Comparative Study of Citizenship Behavior. *The Journal of Politics*, 68(4):783-802.

Bafumi, J. and Herron, M. C. (2010). Leapfrog representation and extremism: A study of american voters and their members in congress. *The American Political Science Review*, 104(3):519-542.

Barabas, J., Jerit, J., Pollock, W., and Rainey, C. (2014). The Question(s) of Political Knowledge. *The American Political Science Review*, 108(4):840-855.

Batista Pereira, F. (2019). Gendered Political Contexts: The Gender Gap in Polit- ical Knowledge. *The Journal of Politics*, 81(4):1480-1493.

Benz, M. and Stutzer, A. (2004). Are Voters Better Informed When They Have a Larger Say in Politics? Evidence for the European Union and Switzerland. *Public Choice*, 119(1-2):31-59.

Berggren, H. M. (2001). Institutional Context and Reduction of the Resource Bias in Political Sophistication. *Political Research Quarterly*, 54(3):531-552.

Blais, A. and Gidengil, E. (1993). Things Are Not Always What They Seem: French-English Differences and the Problem of Measurement Equivalence. *Canadian Journal of Political Science*, 26(3):541-555.

Borjas, G. J. and Sueyoshi, G. T. (1994). A two-stage estimator for probit models with structural group effects. *Journal of Econometrics*, 64:165-182.

Bowler, S. and Donovan, T. (2002). Democracy, Institutions and Attitudes about Citizen Influence on Government. *British Journal of Political Science*, 32(2):371- 390.

Carey, J. M. and Hix, S. (2011). The Electoral Sweet Spot: Low-Magnitude Proportional Electoral Systems. *American Journal of Political Science*, 55(2):383-397.

Castells, M. (1996). *The Rise of the Network Society*. John Wiley and Sons, West Sussex, UK.

Clinton, J., Jackman, S., and Rivers, D. (2004). The Statistical Analysis of Roll Call Data. *The American Political Science Review*, 98(2):355-370.

Costello, A. B. and Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research & Evaluation*, 10(7):1-9.

Dassonneville, R. and McAllister, I. (2018). Gender, Political Knowledge, and Descriptive Representation: The Impact of Long-Term Socialization. *American Journal of Political Science*, 62(2):249-265.

Davidov, E., Dulmer, H., Cieciuch, J., Kuntz, A., Sedig, D., and Schmidt, P. (2018). Explaining Measurement Nonequivalence Using Multilevel Structural Equation Modeling: The Case of Attitudes Toward Citizenship Rights. *Sociological Methods and Research*, 47(4):729-760.

DeBell, M. (2013). Harder Than It Looks: Coding Political Knowledge on the ANES. *Political Analysis*, 21:393-406.

Delli Carpini, M. X. and Keeter, S. (1996). *What Americans Know About Politics and Why It Matters*. Yale University Press, New Haven.

Elff, M. (2009). Political Knowledge in Comparative Perspective: The Problem of Cross-National Equivalence of Measurement. Paper presented at the MPSA 2009 Annual National Conference.

Embretson, S. E. and Reise, S. P. (2000). *Item Response Theory For Psychologists*.

Lawrence Earlbaum Associates, New Jersey.
Fortin-Rittberger, J. (2016). Cross-National Gender Gaps in Political Knowledge: How Much is Due to Context? *Political Research Quarterly*, 69(3):391-402.
Fortin-Rittberger, J. (2019). Political Knowledge: Assessing the Stability of Gen- der Gaps Cross-Nationally. *International Journal of Public Opinion Research*, edz005:1-20.
Fraile, M. (2013). Do information-rich contexts reduce knowledge inequalities? The contextual determinants of political knowledge in Europe. *Acta Politica*, 48(2):119-143.
Gallagher, M. (1991). Proportionality, Disproportionality, and Electoral Systems.
Electoral Studies, 10(1):33-51.
Gibson, J. L. and Caldeira, G. A. (2009). Knowing the Supreme Court? A Re- consideration of Public Ignorance of the High Court. *The Journal of Politics*, 71(2):429-441.
Gordon, S. B. and Segura, G. M. (1997). Cross-National Variation in the Political Sophistication of Individuals: Capability or Choice? *The Journal of Politics*, 59(1):126-147.
Gronlund, K. and Milner, H. (2006). The Determinants of Political Knowledge in Comparative Perspective. *Scandinavian Political Studies*, 29(4):386-406.
Iyengar, S. (1976). Assessing Linguistic Equivalence in Multilingual Surveys. *Com- parative Politics*, 8(4):577-589.
Iyengar, S., Curran, J., Lund, A. B., Salovaara-Moring, I., Hahn, K. S., and Coen,
S. (2010). Cross-National versus Individual-Level Differences in Political Infor- mation: A Media Systems Perspective. *Journal of Elections, Public Opinion, and Parties*, 20(3):291-309.
Iyengar, S., Hahn, K. S., Bonfadelli, H., and Marr, M. (2009). «Dark Areas of Ignorance» Revisited: Comparing International Affairs Knowledge in Switzerland and the United States. *Communication Research*, 36(3):341-358.
Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*. John Wiley and Sons, West Sussex, UK.
Johnson, V. E. and Albert, J. H. (1999). *Ordinal Data Modelling*. Springer, New York.
King, G., Keohane, R. O., and Verba, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton University Press, Princeton.
Knack, S. and Keefer, P. (1997). Does Social Capital Have an Economic Payoff? A Cross-Country Investigation. *The Quarterly Journal of Economics*, 112(4):1251-1288.
Kuklinski, J. H., Quirk, P. J., Jerit, J., and Rich, R. F. (2001). The Political Environment and Citizen Competence. *American Journal of Political Science*, 45(2):410-424.
Lau, R. R. and Redlawsk, D. P. (2001). Advantages and Disadvantages of Cognitive Heuristics in Political Decision Making. *American Journal of Political Science*, 45(4):951-971.
Lewis, J. B. and Linzer, D. A. (2005). Estimating Regression Models in Which the Dependent Variable Is Based on Estimates. *Political Analysis*, 13:345-364.
Lipset, S. M. (1959). Some Social Requisites of Democracy: Economic Development and Political Legitimacy. *The American Political Science Review*, 53(1):69-105.
Lo, J., Proksch, S.-O., and Gschwend, T. (2014). A Common Left-Right Scale for Voters and Parties in Europe. *Political Analysis*, 22:205-223.
Luskin, R. C. (1987). Measuring Political Sophistication. *American Journal of Political Science*, 31(4):856-899.
Luskin, R. C. and Bullock, J. G. (2011). «Don't Know» Means «Don't Know»: DK Responses and the Public's Level of Political Knowledge. *The Journal of Politics*, 73(1):547-557.
Milner, H. (2002). *Civic Literacy: How Informed Citizens Make Democracy Work*.

University Press of New England, Hanover and London.
Mondak, J. J. (1999). Reconsidering the Measurement of Political Knowledge.
Political Analysis, 8(1):57-82.
Mondak, J. J. and Anderson, M. R. (2004). The Knowledge Gap: A Reexamination of Gender-Based Differences in Political Knowledge. *The Journal of Politics*, 66(2):492-512.
Mondak, J. J. and Canache, D. (2004). Knowledge Variables in Cross-National Social Inquiry. *Social Science Quarterly*, 85(3):539-558.
Neuman, W. R. (1981). Differentiation and Integration: Two Dimensions of Political Thinking. *American Journal of Sociology*, 86(6):1236-1268.
Pérez, E. O. (2009). Lost in Translation? Item Validity in Bilingual Political Surveys. *The Journal of Politics*, 71(4):1530-1548.
Prior, M. and Lupia, A. (2008). Money, Time, and Political Knowledge: Distin- guishing Quick Recall and Political Learning Skills. *American Journal of Political Science*, 52(1):169-183.
Przeworski, A. and Teune, H. (1966-1967). Equivalence in Cross-National Research.
Public Opinion Quarterly, 30(4):551-568.
Rainey, C. (2015). Strategic mobilization: Why proportional representation de- creases voter mobilization. *Electoral Studies*, 37:86-98.
Robison, J. (2015). Who Knows? Question Format and Political Knowledge. *In- ternational Journal of Public Opinion Research*, 27(1):1-21.
Schoonvelde, M. (2013). Media freedom and the institutional underpinnings of political knowledge. *Political Science Research and Methods*, (FirstView):1-16.
Sniderman, P. M. and Levendusky, M. S. (2007). An institutional theory of political choice. In Dalton, R. J. and Klingemann, H.-D., editors, *Oxford Handbook of Political Behavior*, chapter Chapter 23, pages 437-456. Oxford.
Tausanovitch, C. and Warshaw, C. (2013). Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities. *The Journal of Politics*, 75(2):330-342.
Toka, G. (2008). Citizen information, election outcomes and good governance. *Electoral Studies*, 27:31-44.
Verba, S. (1993). The Uses of Survey Research in the Study of Comparative Politics: Issues and Strategies. *Historical Social Research*, 18(2):55-103.
Zechmeister, E. J. and Corral, M. (2013). Individual and Contextual Constraints on Ideological Labels in Latin America. *Comparative Political Studies*, 46(6):675-701.
Zeller, R. A. and Carmines, E. G. (1980). *Measurement in the social sciences: The link between theory and data*. Cambridge University Press.