

Comparative Research is Harder Than We Thought: Regional Differences in Experts’ Understanding of Electoral Integrity Questions – Supplementary Material

Bruno Castanho Silva
University of Cologne
bcsilva@wiso.uni-koeln.de

Levente Littvay
Central European University
littvayl@ceu.edu

PEI Methodology and Validity

Here we summarize the methodological and technical information from the PEI 5.5 Codebook, available at <https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/EWYTZ7/HQEZK9&version=2.0>. The Perceptions of Electoral Integrity (PEI) is a dataset where experts are asked to evaluate the electoral integrity of specific elections in countries on which they specialize. In the 5.5 version, which we analyze, it covers 260 elections in 161 countries between 1 July 2012 and 30 June 2017. The criteria for inclusion of a country are that it is a member of the UN, is not a microstate (population below 100,000), had direct elections for the lower house of the legislature and at least one was conducted in the last 30 years.

Regarding the selection of experts, an expert is defined as a political scientist or social scientist in a related discipline who has published academic research or demonstrated knowledge (through participation in conferences, membership in professional academic associations, or being employed in a higher education institution) on the electoral process of a given country. At least forty experts are contacted for each election, with a 50/50 balance between locals and international experts. They are contacted a month after the elections, with three reminders after that. The survey can be taken in three languages: French, Spanish, or English, and the majority chooses the latter.

There have been several tests of internal and external validity of the PEI dataset. Most focus on how experts’ personal characteristics (such as ideology or whether they

are domestic or international) may help predicting their answers (Norris, Wynter and Cameron, 2017); on how characteristics of elections predict disagreements (Martínez i Coma and Ham, 2015), and that it has external validity, being correlated with other indicators of democratic quality (Norris, Frank and Martínez i Coma, 2014). However, none of these tests uses DIF or measurement invariance methods to investigate whether there are differences in question understanding from experts across countries or regions, which is what we apply in this paper.

Descriptive Statistics

Table S1: Descriptive Statistics of the Perceptions of Electoral Integrity Dataset

Region	Countries	Elections	Experts	Response Rates
Africa	46	66	544	22%
Americas	29	40	455	28%
Asia	39	67	726	28%
Europe	39	76	1135	34%
Oceania	9	12	101	23%
East & Southern Africa	21	26	262	26%
West & Central Africa	21	33	228	19%
East Asia & Pacific	23	34	387	28%
South Asia	7	9	119	29%
Eastern Europe	28	62	730	30%
Western Europe	20	34	595	41%
Middle East	9	16	144	23%
North Africa	4	6	41	18%
Americas	29	40	455	28%

Anchoring Vignettes

The PEI dataset includes two sets of three anchoring vignettes each – one set used until 2016, and the other for 2017. The text of Vignettes used (2013-2016) is below. In all, respondents are given a 1-10 response scale where 1 means that electoral integrity is not seriously undermined, and 10 means it is seriously undermined:

- How seriously do you think that electoral integrity is undermined if in STATE A some voters had to wait in long lines to vote?
- How seriously do you think that electoral integrity is undermined in STATE B if the opposition decides to boycott an election, so that the government wins most seats by default?
- How seriously do you think that electoral integrity is undermined in STATE C election results lead to widespread violence throughout the country?

For each of the 49 indicators, we use the Wand et al. (2016) package for R to generate DIF-corrected ranked responses for each individual respondent, with the ‘C’ method. If the model returns an interval (i.e., DIF-corrected responses would be within a given plausible range, say 5–7, and not a single number), we take the mean of that range. Since the **anchors** model cannot deal with missing data, we have to apply listwise deletion for each dimension – for each dimension separately, we remove all observations that have missingness in one of the vignettes or in one of the indicators forming that dimension. The resulting N ’s for each dimension are: Electoral Laws (2224); Procedures (2254); Boundaries (1504); Voter Registration (1701); Party Registration (1930); Media Coverage (1704); Campaign Finance (1636); Voting Process (1060); Vote Count (1726); Voting Results (1765); and Electoral Authorities (2151).

Full Results with Metric and Scalar Invariance

Table S2: Measurement Invariance of the Eleven Dimensions of Electoral Integrity.

Dimension	Metric invariance			Scalar invariance		
	$\Delta\chi^2(df)$	<i>p-value</i>	RMSEA (Δ)	CFI (Δ)	$\Delta\chi^2(df)$	<i>p-value</i> RMSEA (Δ) CFI (Δ)
5 Continents						
Laws	10.274(8)	.246	.025(.025)	.999(.001)	21.147(8)	.007 .040(.015) .993(.006)
Procedures	20.737(12)	.054	.053(.003)	.987(.006)	69.204(12)	<i>p</i> < .001 .062(.009) .972(.015)
Boundaries	6.509(8)	.590	.000(.000)	1.00(.000)	56.144(8)	<i>p</i> < .001 .082(.082) .957(.043)
Voter registration	6.407(8)	.602	.000(.000)	1.00(.000)	22.342(8)	.004 .035(.035) .994(.006)
Party registration	83.399(16)	<i>p</i> < .001	.222(.031)	.752(.051)	231.937(16)	<i>p</i> < .001 .216(.006) .674(.078)
Media coverage	35.102(16)	.004	.169(.027)	.871(.023)	107.266(16)	<i>p</i> < .001 .164(.005) .832(.039)
Campaign finance	16.211(16)	.438	.184(.032)	.835(.026)	278.932(16)	<i>p</i> < .001 .190(.005) .757(.078)
Voting process	69.626(28)	<i>p</i> < .001	.135(.008)	.874(.015)	282.821(28)	<i>p</i> < .001 .142(.007) .830(.044)
Vote count	29.051(16)	.024	.217(.016)	.657(.102)	196.180(16)	<i>p</i> < .001 .215(.002) .532(.125)
Voting results	27.646(12)	.006	.058(.003)	.987(.008)	115.397(12)	<i>p</i> < .001 .083(.025) .960(.028)
Electoral authorities	13.359(12)	.343	.064(.011)	.982(.007)	44.497(12)	<i>p</i> < .001 .066(.003) .970(.012)

9 Regions

Laws	39.141(16)	.001	.077(.077)	.987(.013)	63.752(16)	<i>p</i> < .001 .089(.012) .965(.022)
Procedures	47.254(24)	.003	.061(.000)	.985(.009)	62.372(24)	<i>p</i> < .001 .064(.004) .973(.012)
Boundaries	12.630(16)	.699	.00(.000)	1.00(.000)	63.479(16)	<i>p</i> < .001 .071(.071) .971(.029)
Voter registration	20.131(16)	.214	.037(.037)	.996(.004)	37.415(16)	.002 .057(.020) .982(.014)
Party registration	76.600(32)	<i>p</i> < .001	.224(.048)	.778(.030)	226.59(32)	<i>p</i> < .001 .217(.007) .706(.072)
Media coverage	55.279(32)	.006	.180(.043)	.849(.015)	113.023(32)	<i>p</i> < .001 .170(.010) .808(.041)
Campaign finance	53.01(32)	.011	.205(.045)	.816(.024)	338.76(32)	<i>p</i> < .001 .209(.004) .730(.086)
Voting process	121.10(56)	<i>p</i> < .001	.166(.016)	.838(.014)	312.60(56)	<i>p</i> < .001 .170(.003) .792(.047)
Vote count	386.73(32)	<i>p</i> < .001	.268(.005)	.595(.159)	<i>Model failed to converge</i>	
Voting results	33.036(24)	.103	.062(.017)	.985(.005)	115.738(24)	<i>p</i> < .001 .084(.022) .957(.028)
Electoral authorities	23.090(24)	.514	.077(.022)	.974(.007)	91.110(24)	<i>p</i> < .001 .084(.006) .953(.021)

Notes: $\Delta\chi^2$: difference between the less restricted's and the respective (scalar or metric) invariance model's χ^2 statistics; *df*: difference in the number of free parameters between the two models, and p-values for this difference. RMSEA and CFI from the invariance models; Δ RMSEA and Δ CFI are the difference in RMSEA and CFI from that model to the less restricted one. All fit statistics are calculated with the Satorra and Bentler (2001) correction for robust estimation.

Results without Anchoring Vignettes

Table S3: Measurement Invariance of the Eleven Dimensions of Electoral Integrity.

Dimension	$\Delta\chi^2(df)$	<i>p-value</i>	RMSEA (Δ RMSEA)	CFI (Δ CFI)
5 Continents				
Laws	9.524(8)	.30	.018(.018)	.999(.001)
Procedures	28.716(12)	.004	.072(.015)	.981(.007)
Boundaries	10.681(8)	.221	.029(.029)	.997(.003)
Voter registration	14.425(8)	.071	.039(.039)	.997(.003)
Party registration	64.85(16)	$p < .001$.213(.071)	.555(.037)
Media coverage	115.32(16)	$p < .001$.126(.013)	.888(.029)
Campaign finance	68.80(16)	$p < .001$.110(.017)	.879(.022)
Voting process	183.44(28)	$p < .001$.108(.003)	.731(.046)
Vote count	35.50(16)	.003	.212(.107)	.669(.125)
Voting results	53.91(12)	$p < .001$.096(.028)	.916(.020)
Electoral authorities	44.287(12)	$p < .001$.079(.009)	.976(.011)
9 Regions				
Laws	54.055(16)	$p < .001$.085(.085)	.979(.021)
Procedures	81.29(24)	$p < .001$.086(.000)	.970(.017)
Boundaries	27.358(16)	.038	.056(.056)	.991(.009)
Voter registration	27.551(16)	.036	.050(.050)	.994(.006)
Party registration	139.53(32)	$p < .001$.199(.064)	.622(.009)
Media coverage	114.98(32)	$p < .001$.120(.018)	.895(.024)
Campaign finance	52.02(32)	.014	.092(.023)	.910(.008)
Voting process	144.36(56)	$p < .001$.119(.014)	.693(.014)
Vote count	115.12(32)	$p < .001$.216(.052)	.648(.036)
Voting results	77.59(24)	$p < .001$.105(.046)	.899(.011)
Electoral authorities	72.17(24)	$p < .001$.078(.001)	.975(.014)

Notes: $\Delta\chi^2$: difference between the configural model's and the metric invariance model's χ^2 statistics; df : difference in the number of free parameters between the two models. p -values for this difference in χ^2 are calculated with the Satorra and Bentler (2001) correction for robust estimation. RMSEA and CFI are from the metric invariance model; Δ RMSEA and Δ CFI are the difference in RMSEA and CFI from the configural to the loadings-equality (metric) model.

Table S4: Measurement Invariance with the Alignment Method:
Loadings and Intercepts Invariance by Region.

Indicator		5 continents					9 regions								
		1	2	3	4	5	1	2	3	4	5	6	7	8	9
Laws	Electoral laws were unfair to smaller parties														
	Electoral laws favored the governing party or parties														
	Elections laws restricted citizens' rights				.14										
Proced.	Elections were well managed			-.13											
	Information about voting procedures was widely available					.17									
	Election officials were fair			.34											
	Elections were conducted in accordance with the law														
Bound.	Boundaries discriminated against some parties				-.16										
	Boundaries favored incumbents														
	Boundaries were impartial					-.46									
V. reg.	Some citizens were not listed in the register		.19		.17	-.32				-.27	.23	.17			.20
	The electoral register was inaccurate				-.13										
	Some ineligible electors were registered														
Party reg.	Some opposition candidates were prevented from running														
	Women had equal opportunities to run for office														
	Ethnic and national minorities had equal opportunities to run for office														
	Only top party leaders selected candidates														
	Some parties/candidates were restricted from holding campaign rallies														
Coverage	Newspapers provided balanced election news														
	TV news favored the governing party	-.58		-.26			-.57			.29		.25			.24
	Parties/candidates had fair access to political broadcasts and advertising	.20			-.18		.43								
	Journalists provided fair coverage of the elections														
	Social media were used to expose electoral fraud	-.27		-.31										-.31	
Resources	Parties/candidates had equitable access to public political subsidies														
	Parties/candidates had equitable access to political donations											.25			
	Parties/candidates publish transparent financial accounts			-.18			.22								
	Rich people buy elections			-.17								.50			.21
	Some state resources were improperly used for campaigning											.45			
Voting process	Some voters were threatened with violence at the polls														
	Some fraudulent votes were cast														
	The process of voting was easy											-.31			
	Voters were offered a genuine choice at the ballot box	-.36	-.34									.39			
	Postal ballots were available								-.33	-.53					
	Special voting facilities were available for the disabled						.44								
	National citizens living abroad could vote	.28							-.69			-.58	-.57		
	Some form of internet voting was available													-.72	-.36
Count	Ballot boxes were secure														
	The results were announced without undue delay			-.45	-.45		.36						-.29		
	Votes were counted fairly						.33	.29				.27			
	Int'l election monitors were restricted														
	Domestic election monitors were restricted														
Results	Parties/candidates challenged the results														
	The election led to peaceful protests														
	The election triggered violent protests														
	Any disputes were resolved through legal channels		-.31				.35				.35		.36	.35	
El. Auth.	The election authorities were impartial				-.11										
	The authorities distributed information to citizens														
	The authorities allowed public scrutiny of their performance														
	The election authorities performed well		-.17												-.14
		1	2	3	4	5	1	2	3	4	5	6	7	8	9
		No fill: Invariant loading and intercept			NI int.		NI load.			NI load. and int.					

Notes: Numbers indicate the standardized difference between the non-invariant group intercept and the average intercept between invariant groups. Continents: 1: Africa; 2: Americas; 3: Asia; 4: Europe; 5: Oceania. Regions: 1: East & Southern Africa; 2: West & Central Africa; 3: East Asia & Pacific; 4: South Asia; 5: Eastern Europe; 6: Western Europe; 7: Middle East; 8: North Africa; 9: Americas.

Alternative ways of Including Anchor Vignettes

The example models described in Figures 1 and 2 incorporate those vignettes by using them to create an Anchor (ANC) latent variable which predicts either the latent variable of the electoral integrity dimension (DIM) being measured (Figure 1) or the indicators (Q1–Q4) that compose that dimension (Figure 2). In this case, the two latent variables are uncorrelated. Results of these two models for all dimensions are in Table S5.

Figure 1: Model with Anchor Predicting Latent Variable

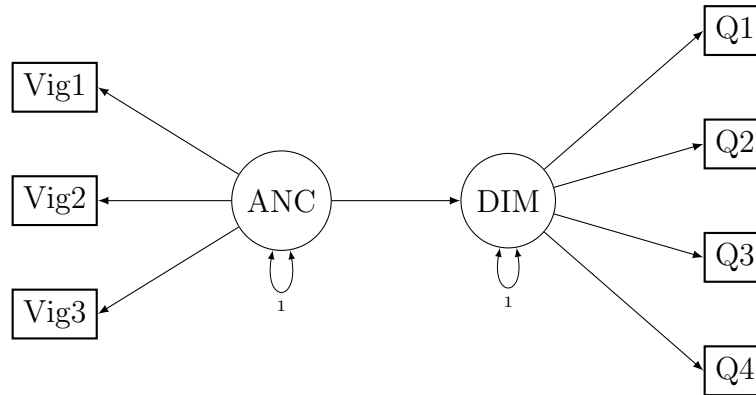


Figure 2: Model with Anchor Predicting Indicators

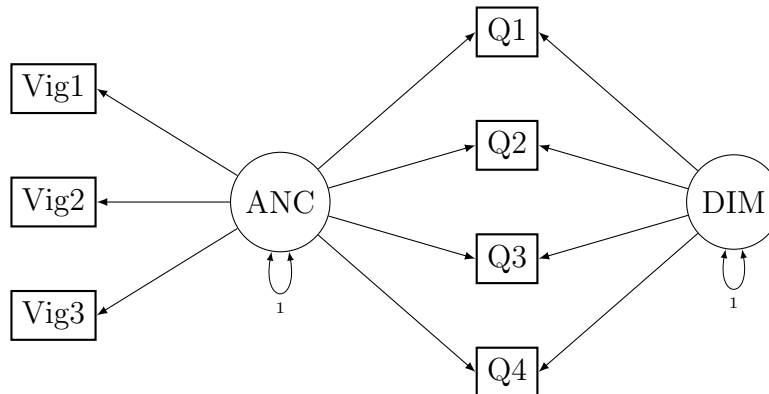


Table S5: Measurement Invariance of the PEI Dimensions with an Anchor

5 Continents				
Category	LV Predicts Factor		LV Predicts Indicators	
	$\Delta\chi^2(df)$	<i>p-value</i>	$\Delta\chi^2(df)$	<i>p-value</i>
Laws	36.589(12)	$p < .001$	38.047(12)	$p < .001$
Procedures	83.339(16)	$p < .001$	81.068(16)	$p < .001$
Boundaries	18.301(12)	.107	19.498(12)	.077
Voter registration	21.585(12)	.042	22.936(12)	.028
Party registration	129.66(20)	$p < .001$	68.647(20)	$p < .001$
Media coverage	104.18(20)	$p < .001$	87.901(20)	$p < .001$
Campaign finance	56.678(20)	$p < .001$	51.944(20)	$p < .001$
Voting Process	127.49(32)	$p < .001$	130.72(32)	$p < .001$
Vote count	50.757(20)	$p < .001$	52.035(20)	$p < .001$
Voting results	85.69(16)	$p < .001$	90.981(16)	$p < .001$
Electoral authorities	106.52(16)	$p < .001$	113.33(16)	$p < .001$
9 Regions				
	$\Delta\chi^2(df)$	<i>p-value</i>	$\Delta\chi^2(df)$	<i>p-value</i>
Laws	91.523(24)	$p < .001$	104.01(24)	$p < .001$
Procedures	148.47(32)	$p < .001$	148.42(32)	$p < .001$
Boundaries	39.173(24)	.026	41.411(24)	.015
Voter registration	41.842(24)	.013	46.015(24)	.004
Party registration	146.69(40)	$p < .001$	124.94(40)	$p < .001$
Media coverage	137.97(40)	$p < .001$	135.91(40)	$p < .001$
Campaign finance	74.995(40)	$p < .001$	69.648(40)	.002
Voting Process	179.23(56)	$p < .001$	123.73(64)	$p < .001$
Vote count	76.796(40)	$p < .001$	56.091(40)	.047
Voting results	90.608(32)	$p < .001$	101.64(32)	$p < .001$
Electoral authorities	115.37(32)	$p < .001$	111.52(32)	$p < .001$

Notes: $\Delta\chi^2$ is the difference between the χ^2 statistic of the configural model and that of the model with factor loadings equality, where *df* is the difference in the number of free parameters between the two models. All models include a latent variable composed of three anchoring Vignette indicators. In *LV Predicts Factor*, the Vignette latent variable predicts the electoral integrity dimension latent variable. In *LV Predicts Indicators*, the Vignette latent variable predicts all individual indicators separately, and is not correlated with the dimension latent variable.

Results in Table S6 follow a two-step approach to incorporate the vignettes, following Bakk and Kuha (2017). In it, first a CFA model only with the vignettes is fit. Then, we extract factor loadings and indicators intercepts of the vignettes, and fix those into the model in Figure 2, leaving the rest as is. In this way, the ‘Anchor’ latent variable is not, in practice, measured also by the indicators themselves.

Table S6: Measurement Invariance of the PEI Dimensions with an Anchor – Two-Step Approach

Category	5 Continents		9 Regions	
	$\Delta\chi^2(df)$	p -value	$\Delta\chi^2(df)$	p -value
Laws	37.974(12)	$p < .001$	104.09(24)	$p < .001$
Procedures	81.02(16)	$p < .001$	148.19(32)	$p < .001$
Boundaries	19.461(12)	.078	41.284(24)	.016
Voter registration	23.084(12)	.027	45.867(24)	.005
Party registration	69.464(20)	$p < .001$	126.32(40)	$p < .001$
Media coverage	88.612(20)	$p < .001$	136.05(40)	$p < .001$
Campaign finance	52.003(20)	$p < .001$	69.646(40)	.003
Voting Process	129.56(32)	$p < .001$	107.57(64)	$p < .001$
Vote count	51.903(20)	$p < .001$	70.16(40)	.002
Voting results	90.938(16)	$p < .001$	100.5(32)	$p < .001$
Electoral authorities	113.6(16)	$p < .001$	112.11(32)	$p < .001$

Notes: $\Delta\chi^2$ is the difference between the χ^2 statistic of the configural model and that of the model with factor loadings equality, where df is the difference in the number of free parameters between the two models. All models include a latent variable composed of three anchoring Vignette indicators, for which loadings have been fixed based on a prior measurement model including only the vignettes.

References

- Bakk, Zsuzsa and Jouni Kuha. 2017. “Two-Step Estimation of Models Between Latent Classes and External Variables.” *Psychometrika* .
URL: <https://doi.org/10.1007/s11336-017-9592-7>

- Martínez i Coma, Ferran and Carolien Van Ham. 2015. “Can experts judge elections? Testing the validity of expert judgments for measuring election integrity.” *European Journal of Political Research* 54(2):305–325.
- Norris, Pippa, Richard W. Frank and Ferran Martínez i Coma. 2014. “Measuring Electoral Integrity around the World: A New Dataset.” *PS: Political Science and Politics* 47(4):789–798.
- Norris, Pippa, Thomas Wynter and Sarah Cameron. 2017. “Corruption and Coercion: The Year in Elections 2017.” *The Electoral Integrity Project* .
URL: <https://static1.squarespace.com/static/58533f31bebafbe99c85dc9b/t/5aa60e298165f533f6462e58/1520832089983/The+Year+in+Elections+2017.pdf>
- Satorra, Albert and Peter M. Bentler. 2001. “A scaled difference chi-square test statistic for moment structure analysis.” *Psychometrika* 66(4):507–514.
- Wand, Jonathan, Gary King and Olivia Lau. 2016. *anchors: Statistical analysis of surveys with anchoring vignettes*. R package version 3.0-8.
URL: <https://CRAN.R-project.org/package=anchors>