# Introduction to the Virtual Issue: Addressing Measurement Comparability in Survey Research.

Sebastián M. Saiegh
Department of Political Science
University of California, San Diego
La Jolla, CA 92093
email: ssaiegh@ucsd.edu

## Measurement Comparability in Survey Research

What should be done when respondents interpret identical questions in vastly different ways? This is a long-standing problem in survey research. Indeed, consideration of measurement invariance is not new to political science (e.g. Aldrich and McKelvey 1977; Brady 1985; Alvarez and Nagler 2004). Problems of interpersonal as well as cross-cultural incomparability in survey research, however, have become more noticeable as more general theories are put to a test in as many different contexts as possible.

Some of the problems of systematic respondent-level bias, or *differential item functioning* (DIF), may be ameliorated by improving the design of survey questions. For example, King et al. (1994) use respondents' assessments of hypothetical individuals described in short vignettes to directly measure response category incomparability. Most scholars, however, do not get to design their own surveys. Instead, they rely on existing large-scale studies such as the Cooperative Congressional Election Study (CCES), the General Social Survey (GSS) and the American National Election Studies (ANES) for the United States; or the World Values Survey, the Comparative Study of Electoral Systems (CSES) and the European Election Study (EES) for cross-national individual level survey data. It is thus critically important to improve the comparability of measurement in survey research when direct indicators of DIF are not available.

## This Virtual Issue

The articles included in this Virtual Issue of *Political Analysis* demonstrate how response incomparability can drastically mislead researchers, and present different strategies and statistical approaches to address this problem. These papers cover a variety of areas in political science (American Politics, Western European Politics, and Latin American Politics) and examine different manifestations of DIF: (1) individual-level respondent bias; (2) biases in scale perception *across* countries; and (3) disjoint groups facing disjoint sets of choices. Nonetheless, they nicely fall into a coherent set of complementary subsets:

- Measurement Invariance
  - Matthew Pietryka and Randall C. MacIntosh, "An Analysis of ANES Items and Their Use in the Construction of Political Knowledge Scales." *Political Analysis* (2013), 21: 407-429.



https://doi.org/10.1017/S104719870001439X Published online by Cambridge University Press

  - Efrén O. Pérez, "The Origins and Implications of Language Effects in Multilingual Surveys: A MIMIC Approach with Application to Latino Political Attitudes." *Political Analysis* (2011), 19:434454.
  - Daniel Stegmueller, "Apples and Oranges? The Problem of Equivalence in Comparative Research." *Political Analysis* (2011), 19 :471487.

- Estimation Strategies

  - Ryan Bakker and Keith T. Poole, "Bayesian Metric Multidimensional Scaling," *Political Analysis* (2013), 21: 125-140.
  - Pablo Barberá, "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* (2015), 23:7691.
  - Thomas König, Moritz Marbach and Moritz Osnabrügge, "Estimating Party Positions across Countries and Time -A Dynamic Latent Variable Model for Manifesto Data." *Political Analysis* (2013), 21: 468-491.

- Empirical Applications

  - James Lo, Sven-Oliver Proksch and Thomas Gschwend, "A Common Left-Right Scale for Voters and Parties in Europe." *Political Analysis* (2014), 22:205223.
  - Sebastián M. Saiegh, "Using Joint Scaling Methods to Study Ideology and Representation: Evidence from Latin America." *Political Analysis* (2015), 23: 363-384.

The first set of papers in the virtual issue is mostly concerned with on the problem of measurement invariance. **Pietryka and MacIntosh (2013)** focus on American National Election Studies (ANES) items used in the construction of political knowledge scales. Using a multi-group confirmatory factor analysis, they demonstrate that these scales are not sufficiently invariant for valid comparisons across a host of theoretically important grouping variables (such as gender, age, education, income, media use, and political participation). A central implication of this finding is that commonly employed knowledge scales can produce misleading estimates of differences in knowledge between sub-populations of interest. To help researchers avoid this problem, the authors identify the items that are most problematic for the construction of measurement invariant knowledge scales. This is an important contribution of the paper, as it provides constructive guidance for future research in this area. Finally, Pietryka and MacIntosh use a vanishing tetrad test (VTT) to assess the nature of the measurement model underlying political knowledge items. Their test suggests it is more appropriate to conceive of these items as effects of a latent variable rather than cause or formative indicators.

The second article, by **Pérez (2011)**, nicely complements the paper on political knowledge scales discussed in the previous paragraph. The author argues that differences of political opinion between English and Spanish interviewees found by previous research may result



https://doi.org/10.1017/S104719870001439X Published online by Cambridge University Press

from two related yet distinct types of language effects: (1) differences in *attitude* and (2) differences in *measures* of attitude. To disentangle these two effects, Pérez uses a statistical framework, known as Multiple Indicators Multiple Causes (MIMIC), and measures of political knowledge, perceptions of Americanism, and Latino attachment from the 2006 Latino National Survey. His findings indicate that the language of interview systematically affect the meaning and interpretation of the survey items, even after controlling for measurement error and individual differences in the latent variable being assessed. In fact, the evidence suggests that language DIF can be a common feature of Latino survey items, even those for factual constructs like political knowledge. The author concludes with a cautionary tale: bilingual items can produce meaningful differences across language; therefore scholars who analyze multilingual survey data should adequately address this problem in their research.

The last article in this group examines the issue of measurement invariance in comparative survey research. In particular, **Stegmueller (2011)** analyzes the existence of perceptual biases caused by country-induced differences in scale usage. Unlike the two papers discussed above, which test invariance using factor analysis, the author introduces a multilevel mixture item response theory (IRT) model with item bias effects to estimate the measurement error. Stegmuller illustrates his approach looking at preferences for social spending among respondents from 12 different countries. Using survey data on the role of government collected by the 1996 International Social Survey Programme, he demonstrates that combining the responses from different countries without correcting for country-item bias would produce biased scores on the latent preference variable. For example, individuals' support for unemployment compensation would be systematically overestimated in Sweden and Switzerland, and systematically underestimated in Australia and New Zealand. More generally, his findings indicate that quantities of interest calculated from estimates based on country-biased items can be grossly misleading. Therefore, in line with the previous articles, Stegmueller warns scholars against the perils of using public opinion measures that lack equivalence across countries.

The next three papers included in the virtual issue do not deal directly with the problem of measurement invariance, but introduce or discuss methodological approaches that can be effectively used to successfully address interpersonal and cross-context incomparability in survey research. For example, one of the most satisfactory approaches to correcting for DIF is the Aldrich-McKelvey scaling procedure (Aldrich and McKelvey 1977). In the fourth paper in the virtual issue, **Bakker and Poole (2013)** show how to apply Bayesian methods to noisy ratio scale distances for both the classical similarities problem as well as the unfolding problem. Their results indicate that Bayesian methods produce essentially the same point estimates as the classical methods, but are superior in that they provide more accurate measures of uncertainty in the data. In their unfolding example, using the 1968 National Election Study candidate feeling thermometers, Bakker and Poole place respondents and politicians on a common ideological space. And, even though they do not explicitly tackle the problem of interpersonal incomparability, they lay the foundations for



https://doi.org/10.1017/S104719870001439X Published online by Cambridge University Press

the Bayesian Aldrich-McKelvey scaling procedure developed in Hare et al. (2015).

The next paper in the virtual issue, by **Barberá (2015)**, proposes a novel solution to the problem of cross-context comparability. Applying scaling techniques such as Aldrich-McKelvey requires some sort of "bridging" information. This limitation usually prevents one from directly comparing policy preferences between disjoint groups responding to disjoint sets of choices. For example, citizens' ideology is usually recovered from voting decisions or self-reported measures from survey data; but legislators' policy preferences are often estimated using observable roll call voting decisions. Barberá shows that using Twitter networks as a source of information about policy positions has the potential to solve this problem. Under the assumption that social networks are homophilic, he develops a Bayesian Spatial Following model that considers ideology as a latent variable, whose value can be inferred by examining which politics actors each user is following. With this model, which is similar in nature to item-response theory models, the author estimates ideal points (with standard errors) for millions of active Twitter users – political elites as well as ordinary citizens, in six different countries (United States, United Kingdom, Spain, Italy, Germany, and the Netherlands). The results indicate that this method successfully classifies political actors and ordinary citizens according to their political orientation, with the locations along the ideological scale being verified by comparisons to positions estimated using roll call voting, party manifestos, and expert surveys.

As Barberá notes, the results for different countries in his study are not directly comparable. The estimation was performed independently, and therefore the resulting dimension does not have a homogeneous scale across countries. The paper by **König, Marbach and Osnabrügge (2013)** explicitly addresses this issue. Producing cross-national measures of party positions is analogous to the problem of estimating comparable preferences across political institutions or over time. A conventional way to address this issue is to use *bridge* actors. König et al. employ a Bayesian factor analytical model and coded manifesto data to estimate the left-right positions of 388 parties competing in 238 elections across twenty-five European Union (EU) member countries and over sixty years. To identify the country bias parameter, the authors exploit the fact that many national parties also compete in transnational elections for seats in the European Parliament (EP) since 1979. For the time-specific bias parameter, they assume that the political party with the largest relative gains in seat shares has no incentive to change its position vis-a-vis other parties in the next election. Their results suggest that estimates without country- and time-specific bias parameters risk serious, systematic bias in about two-thirds of the data. As such, these findings underscore the importance of assessing the comparability of scales in cross-country and longitudinal studies.

The last two papers included in this virtual issue illustrate how joint scaling methods can be used to estimate the ideological location of voters, parties, and politicians with readily available data sources. In their article, **Lo, Proksch and Gschwend (2014)**, estimate these positions using voter self-placements and their placements of political parties



https://doi.org/10.1017/S104719870001439X Published online by Cambridge University Press

on the left-right scale from the 2009 European Election Study (ESS). First, they apply the Aldrich-McKelvey scaling procedure to correct for systematic perceptual biases of survey respondents within countries to place parties and voters on the same national scale. Next, they rescale country-specific estimates into a common cross-national left-right space using European Parliament group memberships as bridging observations. Finally, they generate estimates of uncertainty through a nonparametric bootstrap. Following this procedure, Lo et al. generate voter and party placements that are cross-nationally comparable. Their results indicate that the rescaling yields more accurate estimates of voter and party placements from the surveys on a left-right scale. For example, they examine party and voter locations in the United Kingdom and demonstrate that rescaled estimates significantly improve the model fit in a spatial model of voting with valence. The same is true for a cross-national model of government defection in European elections.

Last, but not least, the virtual issue wraps up with my paper (**Saiegh 2015**). I use some of the aforementioned methods (the Aldrich-McKelvey, and the Bayesian Aldrich-McKelvey scaling procedures) and similar, *bridge*, items from three large-scale surveys to place voters, parties and politicians from different Latin American countries on a common ideological space. First, I employ data from the 2010 *Latinobarómetro* survey to estimate the ideological location of 11,245 respondents in the region. Next, I compare the ideological location of parties and politicians across different countries using the most recent wave of the Universidad de Salamanca's Parliamentary Elites of Latin America (PELA). Finally, I combine some of the PELA surveys with Module 3 of the Comparative Study of Electoral Systems (CSES) to jointly place voters and elected officials in 4 Latin American countries on a common ideological scale. My results reveal that ideology is a significant determinant of vote choice in Latin America. They also suggest that the success of leftist leaders at the polls reflects the views of the voters sustaining their victories. These findings highlight the importance of using a common-space scale to compare disparate populations and call into question a number of recent studies by scholars of Latin American politics who fail to adequately address the issue of measurement invariance.

## Concluding Remarks

As noted above, as more and more political scientists seek to test their theories using survey data obtained form disparate populations, problems of interpersonal as well as cross-cultural incomparability have become more apparent. This virtual issue of *Political Analysis* will hopefully make researchers aware of them. In addition, these contributions should also provide interested scholars with some useful tools to successfully handle these issues.



https://doi.org/10.1017/S104719870001439X Published online by Cambridge University Press

## References

Aldrich, John H., and Richard McKelvey. 1977. "A Method of Scaling with Applications to the 1968 and 1972 Presidential Elections," *American Political Science Review*, 71: 111-130.

Alvarez, R. Michael, and Jonathan Nagler. 2004. "Party System Compactness: Measurement and Consequences," *Political Analysis*, 12: 46-62.

Brady, Henry E. 1985. "The perils of survey research: Inter-personally incomparable responses," *Political Methodology*, 11: 26991.

Hare, Christopher, David A. Armstrong, Ryan Bakker, Royce Carroll, and Keith T. Poole. 2015. "Using Bayesian Aldrich-McKelvey Scaling to Study Citizens' Ideological Preferences and Perceptions," *American Journal of Political Science*, 59: 759-774.

King, Gary, Christopher J.L. Murray, Joshua A. Salomon, and Ajay Tandon. 2004. "Enhancing the Validity and Cross-cultural Comparability of Survey Research," *American Political Science Review* 97: 4.

**About the Author**: Sebastián M. Saiegh is Professor of Political Science at the University of California, San Diego. He has written on statutory policy-making, legislative politics, sovereign borrowing, and electoral forensics. He is currently working on a project studying the relationship between diversity and team performance in the world of professional soccer.

https://doi.org/10.1017/S104719870001439X Published online by Cambridge University Press