

Innovations in Concepts and Measurement for the Study of Peace and Conflict

Christopher J. Fariss* and James Lo^{†‡}

September 17, 2020

Abstract

The development of theoretical concepts is a fundamental part of the scientific process. The critical steps in evaluating concepts is the development, formalization, and validation of measurement models. These steps are challenging because there is no model-free way to measure unobservable or difficult to observe concepts like the level of repression, the number of individuals killed during a civil war, or the perception of members of an out group. The concepts of interest to the community of scholars interested in peace and conflict are often, by definition, difficult to observe. Measurement models offer systematic tools for evaluating the operational procedures designed to generate data translated from such concepts. Each of the articles in this special issue use some form of latent variable model or related measurement model to bring together observable pieces of information and estimate a set of values for the underlying theoretical concept of interest. Though these projects share a common methodological theme, they represent the diversity of substantive interests that span the study of peace and conflict, broadly conceived.

*Assistant Professor, Department of Political Science, University of Michigan, cfariss@umich.educjfariss@umich.edu; cjf0006@gmail.comcjf0006@gmail.com

†Assistant Professor, Department of Political Science, University of Southern California, lojames@usc.edu lojames@usc.edu

‡We would like to thank the many anonymous reviewers who reviewed the entries in this special issues. Extra thanks to the two anonymous super-reviewers who provided excellent suggestions to the author teams included in this special issue. We would like to thank each of the authors who contributed to this special issue for their participation and feedback to each other and to us. Finally, we would like to thank Bertrand Lescher-Nuland and Scott Gates for their help and advice throughout the process of putting together this special issue.

Introduction to special issue

Today, measurement models, latent variable models in particular, are ubiquitous in political science. However, most of this development has occurred in the study of political ideology, especially in the United States and other democratic contexts. These research communities are well integrated because they share both a methodological and substantive focus on ideology and its relationship to the behaviors of different sets of political actors typically in the American political context (e.g., Bond and Messing, 2015; Bonica, 2012; Carroll et al., 2009, 2016; Clinton et al., 2004; Hare et al., 2015; Martin and Quinn, 2002; Martin et al., 2005; Poole, 2005; Poole and Rosenthal, 1991, 1997; Treier and Hillygus, 2009), sometimes in other democratic contexts (Aleman and Saiegh, 2007; Barbera, 2015; Desposato, 2006; Lo et al., 2014, 2016; Rosenthal and Voeten, 2004) and a few examples in authoritarian contexts as well (Blaydes and Linzer, 2008). While latent variable models have been used to study unobservable concepts of interest to scholars of peace and conflict, such as democracy (Pemstein et al., 2010; Reuning et al., 2019; Treier and Jackman, 2008), consolidation of power in non-democratic contexts (Gandhi and Sumner, 2020), military alliances (Benson and Clinton, 2016), political-economic risk (Quinn, 2004), civil-military relations (Kenwick, 2020), respect for human rights (Fariss, 2014, 2019; Schnakenberg and Fariss, 2014), treaty preferences (Fariss, 2018; Lupu, 2016), and UN voting behavior (Voeten, 2000), there are still many other potential conceptual topics of interest that this scholarly community has not yet studied using these measurement and validation tools. The breadth and diversity of the substantive concerns makes coordinating on the most appropriate methodological tools and sharing the latest substantive insights more difficult for this community of scholars. Indeed, as we have already mentioned, much of the innovations for the methods of measurement in political science are grounded in the substantive concerns of American political scholars (e.g., Carroll et al., 2009, 2016; Jesse, 2017; Imai et al., 2017; Lauderdale, 2010). To address these issues, this special issue brings together scholars of Peace and Conflict, broadly conceived, in

order to develop connections on the use of measurement models that measure latent concepts from peace and conflict processes. In this introduction we present each of the contributions to the special issue. We also highlight how each of the author teams address the challenge of validating the new measurements they produce.

Summary of contributions

We are pleased to present 11 articles highlighting different applications of measurement models to peace and conflict research. The special issue brings together a diverse group of scholars interested in making inferences about concepts of interest to peace and conflict scholars that are not directly observable. Each article makes use of some form of latent variable model or related measurement model, which is designed to bring together observable pieces of information and estimate a set of values for the underlying concept of interest. Our hope is that by engaging in a common dialogue about both the conceptual and methodological challenges that each of the articles have faced, that this special issue will help future researchers generate new innovation along both dimensions.

The articles in this special issue span a wide range of applications related to peace and conflict research, and also stand out in highlighting different types of data that scholars working in this area can incorporate into their latent measurement models. We organize these articles into four distinct groups by data type. First, we present a pair of articles using expert scale data to measure respect for civil and political rights (Clay et al., 2020) and identity-based discrimination (Marquardt, 2020). Second, we present another pair of articles that leverage political text data from American Indian constitutions (Cordell et al., 2020) and nuclear proliferation treaty negotiations (Barnum and Lo, 2020). Third, we have a large set of articles that generate latent measurements from monadic features, including data on rebel tactics (Anders, 2020), wartime sexual violence from human rights reports (Krüger and Nordås, 2020), repressive events and human rights data (Fariss et al., 2020), and internet censorship takedown requests (Meserve and Pemstein, 2020). Our forth and

last set of articles focus on using dyadic data measure between parties, including event data on interstate hostility (Terechshenko, 2020), features of bilateral investment treaties (Montal et al., 2020), and data on third-party recognition of self-determination movements (Huddleston, 2020).

As we describe each article below, we wish to note for our readers, that there is no one-size-fits-all solution to developing and accessing a measurement model. At its core, this is the challenge of measurement validation that each group of authors faces. Each group of authors develops application-specific solutions to this challenge but share a common set of features that we highlight in the article summaries below.

Validation terminology and recommendations

Here we define a few construct validity terms that we use in our summary of each article below. There is considerable variation in the usage of the terms across areas of social science, which presents a challenge for scholars using these terms. Moreover, there is no one size fits all approach to measurement validation for any given measurement project, so the selection of validation tools varies from project to project. We therefore devote some space in this introduction to the presentation of definitions of these terms for transparency. We hasten to add though, that individual authors will vary in their usage of these terms. Best practice dictates that these terms are defined clearly, and we do the same here.¹

The term *construct validity* is used broadly to assess some measurement strategy. Here we break construct validity into two types: *translation validity* and *criterion validity*.² Both translation validity (e.g., face and content validity) and criterion validity (e.g., concurrent, convergent/disagreement, predictive/nomological, etc.) are types of construct validity that relate some aspect of a measure, either its operational definition or its empirical content, to a theoretical concept. Translation validity is an assessment of the operational protocol (i.e.,

¹For general advice and discussion about construct validity see Adcock and Collier (2001), Shadish (2010), and Trochim and Donnelly (2008). For specific attention to latent variable models and construct validity see Fariss et al. (2020) and Jackman (2008).

²Sometimes the term *measurement validity* is used instead of *criterion validity*.

data generating process) itself: will the instrument, survey, test, or the specific questions on the test, be effective at eliciting information from the underlying trait of interest? Importantly, no empirical content is necessary for this assessment because it is about how closely we believe the operational protocol maps on to the concept embedded in the theory. Criterion validity is an assessment of the data generated from the operational protocol relative to the concept of interest. This can lead to some confusion or ambiguity in use of terms because, for translation validity, we are considering the link between the operational protocol and theoretical concept, while for criterion validity, we are considering the link between the data itself and theoretical concept.

There are a couple subtypes of translation validity that are commonly that we review here. Though as we note below, there is overlap and sometimes ambiguity between definitions. *Face or content validity* is a specific check to assess whether there will be conceptual (translational) error in the resulting data generated by the operational protocol. Another way of thinking about face/content validity is as a validation technique that links a theoretical concept to the operational protocol used to generate empirical information about that concept. Note here that Adcock and Collier (2001) prefer to not use the term “face validity” because the definition varies from user to user. Instead, they prefer the term content validity. Content validity is simply a check of the operationalization against the relevant content domain for the theory” (Trochim and Donnelly, 2008).

There are many subtypes of criterion validity; we review some of the most relevant types here. *Concurrent validity* is the empirical analog to face validity. It is an empirical assessment that relates the data obtained from the operational protocol to previously obtained or known estimates of the same concept (Adcock and Collier, 2001; Trochim and Donnelly, 2008). In practice, we commonly use concurrent validity with pre-existing categorical information or rank order data in mind. *Convergent validity* is an assessment of the degree to which the estimates from the operationalization are similar to (converges on) estimates from other operationalizations to which the estimates should, in principle, be theoretically similar

(Trochim and Donnelly, 2008). In most latent variable applications, the analysts will compare estimates from one latent variable model against some baseline measurement model (i.e., another set of estimates from a latent variable model or sometimes the categorical information that is included in the latent variable model). *Predictive validity* or (nomological validity) is an assessment of how well a measure is able to predict values of another variable that it conceptually ought to predict, based on prior theory (Adcock and Collier, 2001; Trochim and Donnelly, 2008). The use of predictive validity as a criterion for inference provides a principled tool for assessing measures generated with observational data.³

As we have noted already, there are many other validation terms used to discuss issues of measurement. Broadly, these terms fit into two types of construct validity: *translation validity* and *criterion validity*. In our summaries of each of the 11 articles, we highlight some of the critical and creative validation choices that each of the author teams use in their applied research.

Measurement models for expert-coded data

The first set of articles uses data from expert-generate survey data. Clay et al. (2020) examines data from the Human Rights Measurement Initiative’s practitioner survey, which asks experts to score their own countries on ordinal scales of human rights abuses based on their definitions in international law. The authors tackle a problem that is common to expert data — the comparability of the ordered scales across respondents, a problem also known in the literature as *differential item functioning*. The authors present a latent variable model that corrects for differences across experts. The new latent estimates have good concurrent validity — i.e., the estimates correspond well to the human rights status of well-known cases. The authors also examine the convergent validity of their measures with similar measures of the same concept from the Varieties of Democracy project, focusing particularly on measures

³Note that the use of this term “predictive” does not imply a forecasting or out of sample assessment. See Fariss and Jones (2018) for a discussion of the use of predictive validity from observational or quasi-experimental research design settings.

of torture and extralegal killings.

Marquardt (2020) also tackles a similar set of substantive and methodological issues, but from a different angle. His article focuses on the robustness of regression analyses that use expert-coded data to different forms of measurement error, including scale perception. This article thus similarly focuses on the role that differential item functioning plays in measurement, but with a specific focus on the regression modeling stage. In doing so, Marquardt extends his earlier work using the Varies of Democracy (V-Dem) measurement model to aggregate expert-coded data and focuses substantively on the relationship between identity-based discrimination and the onset of civil conflict. Through a set of simulation exercises, he finds that regression analyses of civil conflict onset which use expert-coded data are relatively robust to expert error, though the level of robustness varies across aggregation techniques.

Measurement models for text data

The next set of articles construct latent measures using text data. Cordell et al. (2020) also use expert-coded data as an important source of validation data, but principally focus on the use of text data in their article. Specifically, they present an original database of American Indian constitutions and use a text-as-data approach to analyze and classify the content of 97 constitutions. Their analysis shows that machine-coded tools used to analyze constitution text closely reproduces the results of expert coding of judiciary functions and guarantees of judicial independence. Stated differently, their measures demonstrate high levels of convergent validity with the human-coded measures generated as part of another study. However, the approach they adopt also has the benefit of being easily scalable to a large number of texts in a transparent and easily reproducible way.

Barnum and Lo (2020) also leverage text data from conference statements about the Nuclear Non-Proliferation Treaty to estimate country-year measurements of preferences along a non-proliferation vs. disarmament dimension. Substantively, these estimates suggest that the gap in preferences between the nuclear and non-nuclear states has not been growing over

time. They test the convergent validity of their estimates against state preference estimates from United Nations General Assembly votes, the predictive validity of their estimates against measures of nuclear latency, and the face (or concurrent) validity of their estimates from an examination of the estimated word parameters.

Measurement models for monadic data

The next set of articles construct latent measures using monadic characteristics of a single geographic entity, usually of a country or country-year. Krüger and Nordås (2020) use such an approach to build latent measures of wartime sexual violence. The main validation of the measures come from the concurrent validity of the source material — the authors use a data set previously published in the *Journal of Peace Research*, which included categorical variables generated using human rights reports from three different organizations as source material, and is specifically designed to measure sexual violence in armed conflict. This latent variable approach to measuring wartime sexual violence outperforms empirical analysis with the original categorical variables. It also offers a principled approach to address issues of under reporting as new sources of information about this type of violence becomes available.

Meserve and Pemstein (2020) also use monadic country-level data in their article, albeit in a very different substantive area. They use takedown request data from online content providers to construct latent measures of internet censorship by country and across time. In addition to the concurrent validity of the raw data, the measures include a test of predictive validity by comparing the latent takedown scores generated from takedown request data with the expert ratings, which are subjective ratings of content regulation. These two measures are largely uncorrelated with one another across regime type but modestly so within regime type. This validity analysis showcases the challenge of the measurement validation while offering a substantively important new measure that the authors used to showcase how liberal democratic states react to violence events. The authors use the new scores to demonstrate how violent opposition induces states to censor digital content and reduce internet freedoms,

even in democratic states like France.

Anders (2020) extends these monadic approaches further to measure latent characteristics at the sub-national national level. Her article uses geo-coded event data on rebel tactics to estimate spatial-temporal variation in territorial control at the level of hexagonal grid cells on a map. These measures are then assessed using convergent validity by comparing them against territorial control measures in Nigeria from the Armed Conflict Location and Event Data project. Anders also demonstrates the predictive validity of her measures by showing that territorial control and deforestation in the aftermath of the peace agreement in Colombia are correlated in the expected direction. Overall, this measurement model provides a principled tool for measuring a fundamental concept for the study of conflict.

Fariss et al. (2020) extend the latent variable model for human rights to generate predictions of one-sided killings of civilians committed by a government. The authors validate their model through a variety of criterion-related validity checks, including their measures correlation with existing measures (convergent validity) and the overall fit of the posterior predictive distributions generated by model parameters (predictive validity). A central innovation of their model is its ability to generate estimates of one-sided killings, even for cases where such events appear likely, but were not recorded in existing data sets. The authors scrutinize these deviant cases as an additional concurrent validity check. This includes countries such as Afghanistan in 1989, Somalia in 1990, and Democratic Republic of Congo in 1994 where one-sided killing is likely but unrecorded.

Measurement models for dyadic data

Our final set of articles use dyadic sources of data. Dyadic data have a long history in the field of international relations, and their prominence in this volume speaks to how data that is particularly prominent in peace research can usefully be incorporated into latent variable models. One particularly important source of dyadic conflict data is event data, which Terechshenko (2020) uses to measure latent interstate hostility at the dyadic level. The article

draws upon two well-known data sets in this area, the Dyadic Militarized Interstate Disputes data and the Phoenix political event data set. Terechshenko tests the concurrent validity of her measures by examining a subset of dyads with well-known enduring rivalries, including China-South Korea, US-Soviet Union/Russia, and France-United Kingdom. She also extends a study on the relationship between international rivalry and terrorist attacks, replacing previous measures of rivalry with her own measure. Her replication successfully reproduces the strong positive effect of rivalry on the count of transnational terrorist attacks, which provides predictive validity evidence that is supportive of the new measure she introduces.

While dyadic data are particularly prominent in studies of conflict, they can also be used to study trade ties. Furthermore, in some cases it may be possible to construct monadic estimates of latent traits using dyadic data. Both of these characteristics appear in the article by Montal et al. (2020), who measure state investment preferences on an investor protection dimension by scaling dyadic characteristics of bilateral investment treaties that different states sign with each other. To help assess the concurrent validity of their estimates, they closely examine a particular case of interest — the finding that the United States has one of the lowest levels of investor protection for foreign investors of all the countries they examine.

The final article by Huddleston (2020) differs from the other dyadic articles in using *directed* dyad data. In particular, Huddleston is interested using third parties' unilateral policy decisions towards self-determination conflicts to generate a continuous measure of latent international sovereignty — the extent to which those aspiring states are tacitly recognized. A key part of Huddleston's theoretical argument is that official recognition and latent international sovereignty are theoretically distinct concepts — empirically he demonstrates the discriminant validity of his measure against the concept of official recognition by exploring U.S. policy towards four legally unrecognized self-determination movements that vary significantly on latent sovereignty. Using his latent measure as a dependent variable, he finds that diplomatic recognition, extant violence, separatist victory, and sour third-party incumbent

relations are positively correlated with his new measure of latent sovereignty.

Concluding remarks

As discussed in brief above, there is no one size fits all solution to measurement. This is the challenge of measurement validation that each group of authors tackled in their contribution to this special issue on measurement for the study of peace and conflict. Each author team creatively and carefully addressed these issues. As with any measurement approach applied to a difficult or impossible to observe concept, relative improvements are always possible. We hope that future researchers will find each of the applications useful for thinking about ways to do just this type of work. In the spirit of innovative and transparent research, we welcome future dialogue that engages with the articles in this issue.⁴ As the editors of this issue, we hope and plan to continue to help facilitate such a dialogue. Overall, we believe that all of the new latent estimates created for the projects described above will be useful to scholars and practitioners interested in using validated measures of these important concepts related to peace and conflict.

⁴Importantly, all of the data, code, and new measures for each project are publicly available on the dataverse archive.

References

- Adcock, R. and D. Collier (2001). Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review* 95(3), 529–546.
- Aleman, E. and S. M. Saiegh (2007). Legislative preferences, political parties, and coalition unity in chile. *Comparative Politics* 39(3), 253–272.
- Anders, T. (2020). Territorial control in civil wars: Theory and measurement using machine learning. *Journal of Peace Research*.
- Barbera, P. (2015). Birds of the same feather tweet together. bayesian ideal point estimation using twitter data. *Political Analysis* 23(1), 76–91.
- Barnum, M. and J. Lo (2020). Is the npt unraveling? evidence from text analysis of review conference statements. *Journal of Peace Research*.
- Benson, B. V. and J. D. Clinton (2016). Assessing the variation of formal military alliances. *Journal of Conflict Resolution* 60(5), 866–898.
- Blaydes, L. and D. A. Linzer (2008). The political economy of women’s support for fundamentalist islam. *World Politics* 60(July), 576–609.
- Bond, R. M. and S. Messing (2015). Quantifying social media’s political space: Estimating ideology from publicly revealed preferences on facebook. *American Political Science Review* 109(1), 62–78.
- Bonica, A. (2012). Ideology and interests in the political marketplace. *American Journal of Political Science* 57(2), 294–311.
- Carroll, R., J. Lewis, J. Lo, K. Poole, and H. Rosenthal (2016). The structure of utility in spatial models of voting. *American Journal of Political Science* 57(4).

- Carroll, R., J. B. Lewis, J. Lo, K. T. Poole, and H. Rosenthal (2009). Comparing nominate and ideal: Points of difference and monte carlo tests. *Legislative Studies Quarterly* 34(4), 555–591.
- Clay, K. C., R. Bakker, A.-M. Brook, J. Daniel W. Hill, and Amanda (2020). Using practitioner surveys to measure human rights: The human rights measurement initiative’s civil & political rights metrics. *Journal of Peace Research*.
- Clinton, J., S. Jackman, and D. Rivers (2004). The statistical analysis of roll call data. *American Political Science Review* 98(2), 355–370.
- Cordell, R., K. S. Gleditsch, F. G. Kern, and L. Saavedra-Lux (2020). Measuring institutional variation across american indian constitutions using automated content analysis. *Journal of Peace Research*.
- Desposato, S. W. (2006). The impact of electoral rules on legislative parties: Lessons from the brazilian senate and chamber of deputies. *Journal of Politics* 68(4), 1018–1030.
- Fariss, C. J. (2014). Respect for human rights has improved over time: Modeling the changing standard of accountability in human rights documents. *American Political Science Review* 108(2), 297–318.
- Fariss, C. J. (2018). Human rights treaty compliance and the changing standard of accountability. *British Journal of Political Science* 48(1), 239–272.
- Fariss, C. J. (2019). Yes, human rights practices are improving over time. *American Political Science Review* 113(3), 868–881.
- Fariss, C. J. and Z. M. Jones (2018). Enhancing validity in observational settings when replication is not possible. *Political Science Research and Methods* 6(2), 365–380.
- Fariss, C. J., M. R. Kenwick, , and K. Reuning (2020). Measurement models. In L. Curini

- and J. Robert J. Franzese (Eds.), *SAGE Handbook of Research Methods in Political Science and International Relations*. SAGE Publications.
- Fariss, C. J., M. R. Kenwick, and K. Reuning (2020). Estimating one-sided-killings from a robust measurement model of human rights. *Journal of Peace Research*.
- Gandhi, J. and J. L. Sumner (2020). Measuring the consolidation of power in non-democracies. *Journal of Politics*.
- Hare, C., D. A. A. II, R. Bakker, R. Carroll, and K. T. Poole (2015). Using bayesian aldrich-mckelvey scaling to study citizens' ideological preferences and perceptions. *American Journal of Political Science* 59(3), 759–774.
- Huddleston, R. J. (2020). Continuous recognition: A latent variable approach to measuring international sovereignty of self-determination movements. *Journal of Peace Research*.
- Imai, K., J. Lo, and J. Olmsted (2017). Fast estimation of ideal points with massive data. *American Political Science Review* 110(4).
- Jackman, S. (2008). Measurement. In J. M. Box-Steffensmeier, H. E. Brady, and D. Collier (Eds.), *The Oxford Handbook of Political Methodology*. Oxford University Press.
- Jesse, S. A. (2017). Don't know responses, personality and the measurement of political knowledge. *Political Science Research and Methods* 5(4), 711–731.
- Kenwick, M. R. (2020). Self-reinforcing civilian control: A measurement-based analysis of civil-military relations. *International Studies Quarterly*.
- Krüger, J. and R. Nordås (2020). A latent variable approach to measuring wartime sexual violence. *Journal of Peace Research*.
- Lauderdale, B. E. (2010). Unpredictable voters in ideal point estimation. *Political Analysis* 18(2), 151–171.

- Lo, J., S.-O. Proksch, and T. Gschwend (2014). A common left-right scale for voters and parties in europe. *Political Analysis* 22(2).
- Lo, J., S.-O. Proksch, and J. Slapin (2016). Ideological clarity in multi-party competition: A new measure and test using election manifestos. *British Journal of Political Science* 46(3).
- Lupu, Y. (2016). Why do states join some universal treaties but not others? an analysis of treaty commitment preferences. *Journal of Conflict Resolution* 60(7), 1219–1250.
- Marquardt, K. L. (2020). How and how much does expert error matter? implications for quantitative peace research. *Journal of Peace Research*.
- Martin, A. D., K. M. Quinn, , and L. Epstein (2005). The median justice on the u.s. supreme court. *North Carolina Law Review* 83, 1275–1321.
- Martin, A. D. and K. M. Quinn (2002). Dynamic ideal point estimation via markov chain monte carlo for the u.s. supreme court, 1953-1999. *Political Analysis* 10(2), 134–153.
- Meserve, S. A. and D. Pemstein (2020). Terrorism and internet censorship. *Journal of Peace Research*.
- Montal, F., C. Potz-Nielsen, and J. L. Sumner (2020). What states want: Estimating ideal points from international investment treaty content. *Journal of Peace Research*.
- Pemstein, D., S. A. Meserve, and J. Melton (2010). Democratic compromise: A latent variable analysis to ten measure of regime type. *Political Analysis* 18(4), 426–449.
- Poole, K. T. (2005). *Spatial Models of Parliamentary Voting*. Cambridge: Cambridge University Press.
- Poole, K. T. and H. Rosenthal (1991). Patterns of congressional voting. *American Journal of Political Science* 35(1), 228–278.

- Poole, K. T. and H. Rosenthal (1997). *A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.
- Quinn, K. M. (2004). Bayesian factor analysis for mixed ordinal and continuous responses. *Political Analysis* 12(4), 338–353.
- Reuning, K., M. R. Kenwick, and C. J. Fariss (2019). Exploring the dynamics of latent variable models. *Political Analysis* 27(4), 503–517.
- Rosenthal, H. and E. Voeten (2004). Analyzing roll calls with perfect spatial voting: France 1946-1958. *American Journal of Political Science* 48(3), 620–632.
- Schnakenberg, K. E. and C. J. Fariss (2014). Dynamic patterns of human rights practices. *Political Science Research and Methods* 2(1), 1–31.
- Shadish, W. R. (2010). Campbell and rubin: A primer and comparison of their approaches to causal inference in field settings. *Psychological Methods* 12(1), 3–17.
- Terechshenko, Z. (2020). Hot under the collar: A latent measure of interstate hostility. *Journal of Peace Research*.
- Treier, S. and D. S. Hillygus (2009). The nature of political ideology in the contemporary electorate. *Public Opinion Quarterly* 73(4), 679–703.
- Treier, S. and S. Jackman (2008). Democracy as a latent variable. *American Journal of Political Science* 52(1), 201–217.
- Trochim, W. M. and J. P. Donnelly (2008). *Research Methods Knowledge Base* (3rd ed.). Mason, OH: Atomic Dog.
- Voeten, E. (2000). Clashes in the assembly. *International Organization* 54(2), 185–215.