

Innovations in concepts and measurement for the study of peace and conflict

Christopher J Fariss 

Department of Political Science, University of Michigan

James Lo

Department of Political Science and International Relations, University of Southern California

Journal of Peace Research
2020, Vol. 57(6) 669–678
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0022343320975200
journals.sagepub.com/home/jpr



Abstract

The observation, measurement, and analysis of violent and contentious processes are essential parts of the scientific study of peace and conflict. However, concepts such as the level of repression, the number of individuals killed during a civil war, or the perception of members of an out-group, are often by definition difficult to observe directly. This is because governments, non-state groups, NGOs, international organizations, monitoring organizations, and other actors are not incentivized to make information about their actions systematically observable to analysts. In this context, latent variable models can play a valuable role by aggregating various behavioral indicators and signals together to help measure latent concepts of interest that would not otherwise be directly observable. Each of the articles in this special issue uses some form of a 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. Each of the articles pays special attention to the processes that make the observation of peace and conflict processes so challenging. As we highlight throughout this introductory article, the unifying framework we present in this special issue is validation. Though the substantive content of each of the articles in this special issue varies, they represent the diversity of substantive interests that span the study of peace and conflict, broadly conceived. Overall, we hope that the special issue becomes a standard reference for scholars interested in developing and validating new measurement models for the study of peace and conflict.

Keywords

measurement models, peace and conflict, scientific process, theoretical concepts

Introduction to the special issue

The observation, measurement, and analysis of violent and contentious processes, such as the level of repression, the number of individuals killed during a civil war, or the perception of members of an out-group, are essential parts of the scientific study of peace and conflict. However, governments, non-state groups, NGOs, international organizations, monitoring organizations, and other actors are not always incentivized to make information about their actions systematically observable to analysts. Lack of access and resource constraints on monitoring resources make analyzing contentious political process all the more difficult for scholars. These issues

are well known to scholars of peace and conflict, who have collectively devoted considerable resources to the creation and curation of datasets that are transparently produced. However, issues of bias and measurement error are ever-present challenges for this research community. Fortunately, latent variable models are capable of integrating and analyzing diverse sources of information together and help address issues of bias and error that arise across different political contexts. In this special issue, a group of peace and conflict scholars present

Corresponding author:
cjf0006@gmail.com

cutting-edge research studies that use latent variable modeling tools to study a wide variety of peace and conflict processes. Though the substantive content of each of the articles in this special issue varies, the unifying framework for this special issue is validation, which we highlight throughout this introductory article.

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 a substantive focus on ideology and its relationship to the behaviors of different sets of political actors typically in the US political context (e.g. Bond & Messing, 2015; Bonica, 2012; Carroll et al., 2009, 2016; Clinton, Jackman & Rivers, 2004; Hare et al., 2015; Martin & Quinn, 2002; Martin, Quinn & Epstein, 2005; Poole, 2005; Poole & Rosenthal, 1991, 1997; Treier & Hillygus, 2009), sometimes in other democratic contexts (Alemán & Saiegh, 2007; Barberá, 2015; Desposato, 2006; Lo, Proksch & Gschwend, 2014; Lo, Proksch & Slapin, 2016; Rosenthal & Voeten, 2004) and a few examples in authoritarian contexts as well (Blaydes & 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, Meserve & Melton, 2010; Reuning, Kenwick & Fariss, 2019; Treier & Jackman, 2008), consolidation of power in non-democratic contexts (Gandhi & Sumner, 2020), military alliances (Benson & Clinton, 2016), political-economic risk (Quinn, 2004), civil-military relations (Kenwick, 2020), respect for human rights (Fariss, 2014, 2019; Schnakenberg & 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, many of the innovations for the methods of measurement in political science are grounded in the substantive concerns of US political scholars (e.g. Carroll et al., 2009, 2016; Jessee, 2017; Imai, Lo & Olmstead, 2017; Lauderdale, 2010). To address these challenges, 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 addresses the challenge of validating the new measurements they produce. Overall, we hope that the special issue becomes a standard reference for scholars interested in developing and validating new measurement models for the study of peace and conflict.

Before discussing the contributions in this special issue, however, we wish to provide some context about the nature of this special issue within the discipline, and the specific challenges that peace and conflict studies scholars face in working in this area. As we described briefly above, work in latent variable modeling in political science has historically been dominated by the study of western democracies – especially in the United States, and particularly in the study of political ideology. While we can think of a number of reasons why this may be the case, in our opinion one reason stands out in particular – the availability of high-quality, well-structured data. More specifically, the US system offers roll-call voting data within the context of a stable two-party system, all registered and recorded in public documents in a single language. These data have relatively low levels of missingness, and each vote is relatively well understood in the sense that it is associated with a written bill. The actors of interest (i.e. legislators) are generally well defined, and the principal conflicts that occur (i.e. elections) take place at regular two-year intervals. The straightforward setup of this type of data then lends itself to analysis with statistical models that are already commonly known in the psychometrics literature, such as the popular two-parameter item response model. The structured setup of the data also facilitates straightforward validation of our latent concepts of interest – for example, we can compare roll-call vote-derived estimates of ideology to similar estimates from experts, or use them to predict political behavior outside the legislative arena such as financial donations from public interest groups, largely because these other measures are also measured within each prescribed electoral cycle.

In contrast, it is painfully obvious that the structured and relatively simple conditions that apply in the study of political ideology simply do not hold in the peace and conflict studies world. Rather than a simple two-party system, peace and conflict studies often have a wide range of different participants that appear and disappear over time in unstable and complex ways. Rather than taking votes on well-structured pieces of legislation with relatively clear meaning, these actors take a variety of

actions (i.e. repertoires of contention and repression), which are in turn driven by multidimensional motives, making observing and measuring the actions associated with underlying motives much more difficult. Conflicts occur at irregular intervals, and the data that are produced have much higher levels of interdependence and endogeneity than data that we see in a typical legislative context; the data are generally much messier and more difficult to work with. This often produces situations where analysis with ‘off the shelf’ latent variable models is often inappropriate, and researchers face the additional hurdle of both developing and validating new models for their work. Validation of these measures becomes particularly challenging, because the unstable and irregular nature of the units being measured makes it difficult to gather auxiliary variables that can be used to help validate those measures. Overall, our discussion above points to what we see as the principal opportunity and the principal challenge facing peace and conflict scholars in working in this area. Given the unstructured nature:

- (1) How do we *find* peace and conflict studies data that are sufficiently structured for us to construct latent variable models?
- (2) Once we have latent variable estimates with data, how do we go about *validating* those measures?

In our discussion of the articles below, we focus on the challenge of how each individual article goes about conducting validation, and discuss common terminology around validation. Validation does not have a one-size-fits-all approach, so we see a significant part of this special issue’s contribution as not only a set of articles focused on studying peace and conflict broadly, but also introducing a set of different ideas on how scholars might try to validate measures of different concepts. For this reason, our introduction to the special issue focuses heavily on validation specifically because we hope that it may be a useful guide to a future researcher asking how they might validate their own latent variable model. Moreover, we encourage readers of articles in this issue to consider them both for their substantive inferences about peace and conflict but also for their creative validation efforts.

Before presenting the articles in this special issue, we wish to first propose some suggestions about where we think the most promising opportunities lie for current and future peace and conflict scholars working with data and latent variable models. The articles in this special issue can largely be grouped by their focus on one of four different data types. These include expert-coded

measures, text data, monadic data (i.e. measures of a single geographic entity, like a country or country-year), and dyadic data. There has been significant work along all four of these data sources, and we believe that there is plenty of future opportunity within each category and also in integrating them. Briefly:

- **Expert-coded data**, like the V-Dem project (Lindberg et al., 2014), are an ideal data source in cases where the observed data are completely unstructured, and the desired latent measure is highly specific. In such cases, asking human beings to rate actors along latent dimensions is a helpful approach because it can impose structure on the data where none previously existed (i.e. measuring an irregularly timed latent variable in country-year increments). There is a significant literature in this direction that attempts to use non-expert coders that are cheaply available on online platforms to do this sort of work, which significantly reduces the amount of overhead involved (e.g. Benoit et al., 2016).
- **Text data** is indisputably the data source experiencing the largest amount of growth in recent years. This is largely because the internet has made such data available online in easily accessible formats. Common sources of text data used include social media posts, constitutions, and speeches. Less common sources of text data include content such as treaty data, diplomatic cables, and political biographies. In this area, we highlight two exciting future directions particularly relevant for peace of conflict research. First, there is significant work in bridging text analysis across languages (e.g. Proksch et al., 2019), which has the potential to extent text analysis across linguistic boundaries. Second, with computational resources becoming cheaper we are now seeing the extension of these methods into the domain of image and video analysis (e.g. Sobolev et al., 2020; Xi et al., 2020).
- **Monadic data**: This is likely the most common data source in peace and conflict research, due to the importance of the country-year as a unit of conceptual and analytical interest. However, there are significant opportunities for advances in monadic measurements at smaller levels of aggregation (i.e. subnational actors, shorter time units, and smaller geographies). An excellent example of such research is Baskaran, Min & Uppal (2015), who measure electricity provision in India at the

state level using satellite imagery. And there are many new and ongoing projects that are generating subnational data that are ripe for use with latent variable models (e.g. Cordell et al., 2019; Donnay et al., 2019; Raleigh et al., 2010; Zhukov, Davenport & Kostyuk, 2019).

- **Dyadic data:** Dyadic datasets are well known to the peace and conflict studies community and will continue to be prominent in the field. However, the peace and conflict community has increasingly moved in the direction of modeling entire networks rather than simply dyads (e.g. Cranmer, Desmarais & Menninga, 2012; Gallop, Dorff & Minhas, forthcoming).

Our suggestions above fit well with the set of contributions we describe next. Ultimately, we hope the contributions in this special issue along with the additional sources we have cited above will be useful for research for future scholars who wish to develop and validate novel measurement models that capture concepts related to peace and conflict.

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, 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 & 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 & Nordås, 2020), repressive events and human rights data (Fariss, Kenwick & Reuning, 2020b), and internet censorship takedown requests (Meserve & Pemstein, 2020). Our fourth and last set of articles focus on using dyadic data measures between parties, including event data on interstate hostility (Terechshenko, 2020), features of bilateral investment treaties (Montal, Potz-Nielsen & Sumner, 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 they 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 and criterion validity are types of construct validity that relate some aspect of a measure, either its operational definition or its empirical

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

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

content, to a theoretical concept.³ Translation validity is an assessment of the link between the operational protocol (i.e. data generating process) and the theoretical concept of interest: a translation validity assessment asks if the instrument, survey, test, or the specific questions on the test will be effective at generating information that is consistent with the underlying concept of interest. Importantly, no data are necessary for this assessment because it is about how closely we believe the operational protocol maps on to the concept embedded in the theory. As a simple analogy, does a test designed to measure a student's knowledge about a subject relate to the knowledge domain? The teacher will ask this question prior to holding the test, which then generates the student response data. Criterion validity, on the other hand, is an assessment of data generated from the operational protocol relative to the concept of interest. The use of these related terms 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 (test and concept), while for criterion validity, we are considering the link between the data themselves and theoretical concept (data and concept).

There are many subtypes of translation validity that are commonly used. We review a couple of the most commonly used types 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 asks if the test itself or a specific question on the test are related to the theoretical concept of interest. Note here that Adcock & 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 & Donnelly, 2008), that is, the question on the test or some aspect of the presumed data generating process.

Criterion validity is about the data produced by the test or the data generating process. 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 & Collier, 2001; Trochim & Donnelly, 2008). In practice, we commonly use concurrent validity with pre-existing categorical information or rank order data in mind about a few specific cases. Convergent validity is an assessment of the degree to which the estimates from the operationalization are similar to estimates from other operationalizations to which the estimates should, in principle, be theoretically similar (Trochim & 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 & Collier, 2001; Trochim & Donnelly, 2008). The use of predictive validity as a criterion for inference provides a principled tool for assessing measures generated with observational data.⁴

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 uses in their applied research.

Measurement models for expert-coded data

The first set of articles use data from expert-generated survey data. Clay et al. (2020) examine 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 – that is, the estimates correspond well to the human rights status

³ Common subtypes of translation validity include face and content. Common subtypes of criterion validity include concurrent, convergent/disagreement, predictive/nomological.

⁴ Note that the use of this term 'predictive' does not imply a forecasting or out of sample assessment. See Fariss & Jones (2018) for a discussion of the use of predictive validity from observational or quasi-experimental research design settings.

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 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 Varieties 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 reproduce 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 & 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 & Nordås (2020) use such an approach to build latent measures of wartime sexual violence. The main validation of the measures comes from the concurrent validity of the source material – the authors use a dataset 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 (Cohen & Nordås, 2014). 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 underreporting as new sources of information about this type of violence become available.

Meserve & 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 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 subnational 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, Kenwick & Reuning (2020b) 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 datasets. 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 are 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 datasets in this area, the Dyadic Militarized Interstate Disputes data and the Phoenix political event dataset. Terechshenko tests the concurrent validity of her measures by examining a subset of dyads with well-known enduring rivalries, including China–South Korea, USA–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, Potz-Nielsen & Sumner (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 in 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 US 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 readers will discover in this collection, each group of authors tackle the challenge of measurement validation head on in their contributions to this special issue. In doing so, they present new estimates of important substantive concepts, while generating them with more precision and more attention paid to the sources and consequences of measurement error.

With a focus on expert coded data, we learn about how practitioner-survey generated expert data reveal new patterns of intensity and within country variation of human rights abuse (Clay et al., 2020), and how the relationships between different models that incorporate expert-coded identity-based discrimination questions predict the onset of civil conflict (Marquardt, 2020). We also learn about variation in judicial institutions from a new digital corpus of American Indian constitutions (Cordell et al., 2020) and differences between nuclear and non-nuclear states using text statements

about the Nuclear Non-Proliferation Treaty (Barnum & Lo, 2020). Both of these articles use text-as-data models that are widely applicable to other under-measured historical and contemporary institutional contexts. We also learn about new descriptive patterns of censorship (Meserve & Pemstein, 2020), territorial control (Anders, 2020), and repression and one-sided killings (Fariss, Kenwick & Reuning, 2020) using monadic data, and interstate hostility (Terechshenko, 2020), investor protection (Montal, Potz-Nielsen & Sumner, 2020), and international sovereignty of self-determination movements (Huddleston, 2020), using dyadic data. All of these articles offer novel innovations to standard measurement models in addition to generating new estimates of important substantive concepts.


Each of these article teams creatively and carefully addresses the questions we raise in this introduction. Where do we go from here? In the spirit of innovative and transparent research, we welcome future dialogue that engages with the articles in this special issue. As the editors of this issue, we hope and plan to continue to help facilitate such a dialogue. First, we hope that future researchers will find this special issue useful for thinking about ways to apply and validate measurement models in the study of peace and conflict. Second, we hope that readers will use and build on the new latent variable models and the estimates derived from them.⁵ Third, we hope that future researchers will build more descriptive and valid inferences about peace and conflict processes using the tools and knowledge generated by the authors of this special issue. Overall, this special issue provides many new insights and methodological tools that researchers may build upon.

Acknowledgements

We would like to thank the many anonymous reviewers who reviewed the entries in this special issue. 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 all 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.

⁵ Importantly, all of the data, code, and new measures for each article in this special collection are publicly available on JPR's replication page and on the dataverse archive.

ORCID iD

Christopher J Fariss  <https://orcid.org/0000-0001-9837-186X>

References

- Adcock, Robert & David Collier (2001) Measurement validity: A shared standard for qualitative and quantitative research. *American Political Science Review* 95(3): 529–546.
- Alemán, Eduardo & Sebastián M Saiegh (2007) Legislative preferences, political parties, and coalition unity in Chile. *Comparative Politics* 39(3): 253–272.
- Anders, Therese (2020) Territorial control in civil wars: Theory and measurement using machine learning. *Journal of Peace Research* 57(6): 701–714.
- Barberá, Pablo (2015) Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis* 23(1): 76–91.
- Barnum, Miriam & James Lo (2020) Is the NPT unraveling? Evidence from text analysis of review conference statements. *Journal of Peace Research* 57(6): 740–751.
- Baskaran, Thushyanthan; Brian Min & Yogesh Uppal (2015) Election cycles and electricity provision: Evidence from a quasi-experiment with Indian special elections. *Journal of Public Economics* 126(June): 64–73.
- Benoit, Kenneth; Drew Conway, Benjamin E Lauderdale, Michael Laver & Slava Mikhaylov (2016) Crowd-sourced text analysis: Reproducible and agile production of political data. *American Political Science Review* 110(2): 278–295.
- Benson, Brett V & Joshua D Clinton (2016) Assessing the variation of formal military alliances. *Journal of Conflict Resolution* 60(5): 866–898.
- Blaydes, Lisa & Drew A Linzer (2008) The political economy of women's support for fundamentalist Islam. *World Politics* 60(July): 576–609.
- Bond, Robert M & Solomon 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, Adam (2012) Ideology and interests in the political marketplace. *American Journal of Political Science* 57(2): 294–311.
- Carroll, Royce; Jeffrey B Lewis, James Lo, Keith T Poole & Howard Rosenthal (2009) Comparing NOMINATE and IDEAL: Points of difference and Monte Carlo tests. *Legislative Studies Quarterly* 34(4): 555–591.
- Carroll, Royce; Jeffrey B Lewis, James Lo, Keith T Poole & Howard Rosenthal (2016) The structure of utility in spatial models of voting. *American Journal of Political Science* 57(4): 1008–1028.
- Clay, K Chad; Ryan Bakker, Anne-Marie Brook, Daniel W Hill Jr & Amanda Murdie (2020) Using practitioner surveys to measure human rights: The Human Rights Measurement Initiative's civil and political rights metrics. *Journal of Peace Research* 57(6): 715–727.

- Clinton, Joshua; Simon Jackman & Douglas Rivers (2004) The statistical analysis of roll call data. *American Political Science Review* 98(2): 355–370.
- Cohen, Dara K & Ragnhild Nordås (2014) Sexual violence in armed conflict: Introducing the SVAC dataset, 1989–2009. *Journal of Peace Research* 51(3): 418–428.
- Cordell, Rebecca; K Chad Clay, Christopher J Fariss, Reed M Wood & Thorin M Wright (2019) Changing standards or political whim? Evaluating changes in the content of the US State Department human rights reports. *Journal of Human Rights* 19(1). <https://doi.org/10.1080/14754835.2019.1671175>.
- Cordell, Rebecca; Kristian Skrede Gleditsch, Florian G Kern & Laura Saavedra-Lux (2020) Measuring institutional variation across American Indian constitutions using automated content analysis. *Journal of Peace Research* 57(6): 777–788.
- Cranmer, Skyler J; Bruce A Desmarais & Elizabeth J Menninga (2012) Complex dependencies in the alliance network. *Conflict Management and Peace Science* 29(3): 279–313.
- Desposato, Scott 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.
- Donnay, Karsten; Eric T Dunford, Erin C McGrath, David Backer & David E Cunningham (2019) Integrating conflict event data. *Journal of Conflict Resolution* 63(5): 1337–1364.
- Fariss, Christopher 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, Christopher J (2018) Human rights treaty compliance and the changing standard of accountability. *British Journal of Political Science* 48(1): 239–272.
- Fariss, Christopher J (2019) Yes, human rights practices are improving over time. *American Political Science Review* 113(3): 868–881.
- Fariss, Christopher J & Zachary M Jones (2018) Enhancing validity in observational settings when replication is not possible. *Political Science Research and Methods* 6(2): 365–380.
- Fariss, Christopher J; Michael R Kenwick & Kevin Reuning (2020a) Measurement models. In: Luigi Curini & Robert Franzese (eds) *SAGE Handbook of Research Methods in Political Science and International Relations*. Thousand Oaks, CA: Sage.
- Fariss, Christopher J; Michael R Kenwick & Kevin Reuning (2020b) Estimating one-sided-killings from a robust measurement model of human rights. *Journal of Peace Research* 57(6): 801–814.
- Gallop, Max; Cassy Dorff & Shahryar Minhas (forthcoming) [W]hat lies beneath: Using latent networks to improve spatial predictions of political violence. *International Studies Quarterly*.
- Gandhi, Jennifer & Jane Lawrence Sumner (2020) Measuring the consolidation of power in non-democracies. *Journal of Politics* 82(4): 1545–1558.
- Hare, Christopher; David A Armstrong II, Ryan Bakker, Royce Carroll & Keith 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 Joseph (2020) Continuous recognition: A latent variable approach to measuring international sovereignty of self-determination movements. *Journal of Peace Research* 57(6): 789–800.
- Imai, Kosuke; James Lo & Jonathan Olmsted (2017) Fast estimation of ideal points with massive data. *American Political Science Review* 110(4): 631–656.
- Jackman, Simon (2008) Measurement. In: Janet M Box-Steffensmeier, Henry E Brady & David Collier (eds) *The Oxford Handbook of Political Methodology*. Oxford: Oxford University Press, 119–151.
- Jessee, Stephen A (2017) Don't know responses, personality and the measurement of political knowledge. *Political Science Research and Methods* 5(4): 711–731.
- Kenwick, Michael R (2020) Self-reinforcing civilian control: A measurement-based analysis of civil-military relations. *International Studies Quarterly* 64(1): 71–84.
- Krüger, Jule & Ragnhild Nordås (2020) A latent variable approach to measuring wartime sexual violence. *Journal of Peace Research* 57(6): 728–739.
- Lauderdale, Benjamin E (2010) Unpredictable voters in ideal point estimation. *Political Analysis* 18(2): 151–171.
- Lindberg, Staffan I; Michael Coppedge, John Gerring & Jan Teorell (2014) V-Dem: A new way to measure democracy. *Journal of Democracy* 25(3): 159–169.
- Lo, James; Sven-Oliver Proksch & Thomas Gschwend (2014) A common left–right scale for voters and parties in Europe. *Political Analysis* 22(2): 205–223.
- Lo, James; Sven-Oliver Proksch & Jonathan B Slapin (2016) Ideological clarity in multi-party competition: A new measure and test using election manifestos. *British Journal of Political Science* 46(3): 591–610.
- Lupu, Yonatan (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, Kyle L (2020) How and how much does expert error matter? Implications for quantitative peace research. *Journal of Peace Research* 57(6): 692–700.
- Martin, Andrew D & Kevin M Quinn (2002) Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999. *Political Analysis* 10(2): 134–153.
- Martin, Andrew D; Kevin M Quinn & Lee Epstein (2005) The median justice on the US Supreme Court. *North Carolina Law Review* 83(5): 1275–1321.

- Meserve, Stephen A & Daniel Pemstein (2020) Terrorism and internet censorship. *Journal of Peace Research* 57(6): 752–763.
- Montal, Florencia; Carly Potz-Nielsen & Jane Lawrence Sumner (2020) What states want: Estimating ideal points from international investment treaty content. *Journal of Peace Research* 57(6): 679–691.
- Pemstein, Daniel; Stephen A Meserve & James Douglas Melton (2010) Democratic compromise: A latent variable analysis of ten measures of regime type. *Political Analysis* 18(4): 426–449.
- Poole, Keith T (2005) *Spatial Models of Parliamentary Voting*. Cambridge: Cambridge University Press.
- Poole, Keith T & Howard Rosenthal (1991) Patterns of congressional voting. *American Journal of Political Science* 35(1): 228–278.
- Poole, Keith T & Howard Rosenthal (1997) *A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.
- Proksch, Sven-Oliver; Will Lowe, Jens Wäckerle & Stuart Soroka (2019) Multilingual sentiment analysis: A new approach to measuring conflict in legislative speeches. *Legislative Studies Quarterly* 44(1): 97–131.
- Quinn, Kevin M (2004) Bayesian factor analysis for mixed ordinal and continuous responses. *Political Analysis* 12(4): 338–353.
- Raleigh, Clionadh; Andrew Linke, Håvard Hegre & Joakim Karlsen (2010) Introducing ACLED: An armed conflict location and event dataset. Special data feature. *Journal of Peace Research* 47(5): 651–660.
- Reuning, Kevin; Michael R Kenwick & Christopher J Fariss (2019) Exploring the dynamics of latent variable models. *Political Analysis* 27(4): 503–517.
- Rosenthal, Howard & Erik Voeten (2004) Analyzing roll calls with perfect spatial voting: France 1946–1958. *American Journal of Political Science* 48(3): 620–632.
- Schnakenberg, Keith E & Christopher J Fariss (2014) Dynamic patterns of human rights practices. *Political Science Research and Methods* 2(1): 1–31.
- Shadish, William R (2010) Campbell and Rubin: A primer and comparison of their approaches to causal inference in field settings. *Psychological Methods* 12(1): 3–17.
- Sobolev, Anton; M Keith Chen, Jungseock Joo & Zachary C Steinert-Threlkeld (2020) News and social media accurately measure protest size variation. *American Political Science Review* 114(4): 1343–1351.
- Terechshenko, Zhanna (2020) Hot under the collar: A latent measure of interstate hostility. *Journal of Peace Research* 57(6): 764–776.
- Treier, Shawn & D Sunshine Hillygus (2009) The nature of political ideology in the contemporary electorate. *Public Opinion Quarterly* 73(4): 679–703.
- Treier, Shawn & Simon Jackman (2008) Democracy as a latent variable. *American Journal of Political Science* 52(1): 201–217.
- Trochim, William M & James P Donnelly (2008) *The Research Methods Knowledge Base*, 3rd edition. Mason, OH: Atomic Dog.
- Voeten, Erik (2000) Clashes in the assembly. *International Organization* 54(2): 185–215.
- Xi, Nan; Ma Di, Marcus Liou, Zachary C Steinert-Threlkeld, Jason Anastasopoulos & Jungseock Joo (2020) Understanding the political ideology of legislators from social media images. *Proceedings of the International AAAI Conference on Web and Social Media* 14: 726–737.
- Zhukov, Yuri M; Christian Davenport & Nadiya Kostyuk (2019) Introducing xSub: A new portal for cross-national data on subnational violence. *Journal of Peace Research* 56(4): 604–614.

CHRISTOPHER J FARISS, b. 1981, PhD in Political Science (University of California, San Diego, 2013); Assistant Professor, Department of Political Science and Faculty Associate, Center for Political Studies, Institute for Social Research, University of Michigan, (2016–); research interests: human rights, repression, political violence, foreign policy, social media, political behavior, research design, and measurement.

JAMES LO, b. 1978, PhD in Political Science (University of California, Los Angeles, 2010); Assistant Professor, Department of Political Science and International Relations, University of Southern California; research interests: measurement, American politics, European politics, political polarization.