18

How Have Interpretivism and Big Data Changed the Landscape for Measurement Validation?

Jason Seawright

MEASUREMENT VALIDITY AND VALIDATION

David Collier's work on measurement validity and methods of validation establishes a baseline understanding that still largely describes the state of theory and practice. Scholars have developed some new ideas, and real-world conditions have changed in ways that shift the picture in subtle but important ways, but the road map that Collier's work set up largely remains a useful guide to the landscape.

In work with Robert Adcock (Adcock and Collier 2001),[1] Collier built a conceptual apparatus for understanding the measurement process as a sequence of moves from conceptual systematization, operational definition as an actual plan for measurement, and then creation of an indicator as a series of measurements connected with cases. Each step involves potential pitfalls, with measurement validity relating especially to the connections among systematized concepts, operational definitions, and indicators. Adcock and Collier further provide a typology of tools for exploring the conceptual value and empirical adequacy of different measurement schemes. Content validation involves checking that the indicator includes measurement elements that cover the full range of ideas included in the systematized concept and nothing extraneous. Convergent/discriminant validation involves checking that alternative indicators of the same concept are related in appropriate ways, whereas indicators of different concepts are empirically separable from each other. Finally, nomological/construct validation involves testing measurement schemes by seeing how well they perform within the context of a theoretical model.

Building on this work, Seawright and Collier (2014) offer a somewhat more elaborate typology of intellectual traditions within the domain of

[1] See Chapter 16, this volume.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 07:40:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.028

measurement validation.[2] Seawright and Collier differentiate approaches in terms of the ways that scholars justify their knowledge claims about measurement, finding four broad categories. First, a range of scholars draw on ideas about the internal structure of an indicator and the relative ordering of possible responses to justify some analytic techniques, generating a body of work Seawright and Collier call the *levels of measurement tradition*. Second, scholars devise (more or less rigorous) models of the underlying measurement process to motivate analytical models that test and improve measurement properties, a body of work the article names the *structural equation modeling with latent variables tradition*. Third, some scholars draw on a wide range of (quantitative and qualitative) techniques based on no really rigorous fundamental justification beyond the past experience of themselves and other scholars that those techniques are helpful in solving measurement problems. While any technique can be used in this way, some are typically used like this; Seawright and Collier refer to work in this vein as the *pragmatic measurement tradition*. Finally, some work is based in a deep collection of evidence about one or a few cases – so much evidence that there is, in a sense, an overflow of detail beyond the requirements of any given indicator such that each indicator can be checked against the surplus details available within the case for correspondence. This tradition is referred to as the *case-based tradition*.

In combination, these two chapters continue to provide a substantially useful guide to the contemporary landscape of practice in terms of measurement methods. Scholars continue to use most of the same tools that Collier and his collaborators characterized, with most of the same justifications. However, certain subtle but important changes in the landscape are noticeable, owing to changes in the technology available to scholars, the methods in wide use, and the prominence of some intellectual currents within social science. This chapter explores three ways in which measurement validation in political science and related fields has evolved in subsequent years. The surge and intellectual flourishing of interpretive/ethnographic methods in political science has offered some new perspectives related to measurement validation, but also a suite of fresh approaches to validation research at the individual level, especially for contemporary questions.

Recent years have seen an explosion of hype around the slogan of "big data," followed by a fragmentation of that slogan into much smaller but perhaps more precise research discourses. Two of those streams of thought are of relevance here. As digital data storage and distribution become the norm, researchers have become accustomed to working with data sets whose dimensionality is vastly larger than was the norm (or, perhaps, was even achievable) when Collier's measurement articles were written. In particular, it is trivially easy to obtain data sets that include far more different

[2] See Chapter 17, this volume.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 07:40:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.028

variables offering competing measures of a concept than existed in the 1980s or 1990s. This proliferation of very complex data sets pushes scholars almost necessarily away from in-depth knowledge of the data-generation process behind specific indicators and toward measurement validation tools that ask and answer more general questions.

Finally, scholars within political science have embraced text-as-data methods as a tool for engaging in measurement debates. The formal mathematical properties of these techniques often closely resemble factor analysis and thus fit neatly within the concepts of convergent/discriminant validation and structural equation modeling with latent variables from Collier's work on validation. However, because these newer methods use text, rather than simply numbers, as the input for analysis, scholars have found mixed-method approaches all but irresistible. In practice, the boundaries between the concepts mentioned earlier and the case-based measurement tradition may be eroding in this domain.

INTERPRETIVE PERSPECTIVES AND ETHNOGRAPHIC METHODS

The last decades have seen a major increase in the organization, visibility, and scope of interpretive perspectives and the use of ethnographic research methods in political science generally. This shift has had important consequences for the intellectual landscape connected with measurement and measurement validity. Here, I will highlight two major changes resulting from the surge of interpretive work: renewed and expanded attention to considerations of how power relations, social positionality, and ethical issues interact with questions of validity; and a wonderfully enriched hands-on methodological literature related to ethnographic measurement approaches.

One prominent example is Lee Ann Fujii's influential work on qualitative interview methods, reflected most notably in her book *Interviewing in Social Science Research: A Relational Approach*. Fujii's (2017) book discusses interview technique and analysis with an interest in avoiding blunders, shallowness, and misinterpretations that clearly resonates with long-standing qualitative measurement validity concerns. Yet the book adds several notable new perspectives, two of which I will highlight here.

First, there is the obvious focus on in-person collection of individual-level data from living people, an important contrast from a qualitative and case-based measurement literature that has concentrated on historical measurement, often at the country level. Fujii shows how moving to the individual level opens up space for measurement debates about meaning and identity (Fujii 2017: 9–11), questions that have often been treated as fuzzy or inaccessible to scientific measurement.

Second, Fujii adds the distinctive concerns of interpretive methodology with social positionality (15–22), the ethical treatment of research participants (6–7, 22–28), and a full intellectual engagement with the implications of

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 07:40:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.028

respect for the autonomy and strategic goals of the people being interviewed (12–15). These concerns deepen and, in some ways, reposition considerations regarding validity. Much of the existing literature on validity takes the goal of the measurement process as a given, once the scholar's conceptual choices have been made, but Fujii's work (like that of other interpretivists) reminds us that the outcome of measurement is and necessarily must be a product of the agendas of both the scholar and the person or people being interviewed. Those people have goals of presenting their activities, political movements, and societies in certain ways, and they also have their own social theories that they seek to advocate. The way interview participants present information in pursuit of those goals can distort information and introduce bias, as one might frame such considerations in much of the existing measurement literature. However, as Fujii compellingly argues, these interactions are themselves part of the social content we wish to study, and fluidly readjusting our conceptual framework to incorporate the way we see our interlocutors engaging with us allows us to better fit our social science measurement apparatus with our understanding of people as co-creators of social meaning.

For example, Fujii discusses a series of interviews she carried out with people about violence they experienced during Rwanda's genocide. She originally focused her research on the period of time through the end of 1994, when Rwanda's war and genocide concluded, but she recounts an interview in which a respondent effectively persuaded her to change this measurement strategy:

> I asked one woman what she remembered from 1994. She said she recalled no violence at that time. Incredulous, I reworded the question slightly but stuck to the same time frame. She persisted in stating that she recalled nothing out of the ordinary from that period. I finally let go of that time period and asked about any violence she recalled. It was only then that she told me what happened to her in 1996. In that year, those who had been active in the genocide and had fled to neighboring Zaire (present day Democratic Republic of Congo) began launching attacks from across the border. The Rwandan government responded with counterattacks, and during this war the woman lost her husband and two of her children. Once I realized that my time frame did not have the same significance for everyone, I rephrased my opening question. (63)

From one perspective, one might ask whether it is literally correct that there was no violence in the area this woman lived during 1994, or whether this response entails measurement error. Much of the existing literature works from this perspective, which is of course important. Alternatively, one might seek to understand the woman's goal in downplaying violence from 1994, which in this case was to emphasize the salience of violence later on and to persuade her interlocutor to understand that there was not in fact a fixed endpoint to the violence in Rwanda. That intention in itself is meaningful social data – whether or not the researcher is persuaded, as Fujii was, that it would be good research practice to follow the woman's advice and change measurement strategies.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 07:40:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.028

This pivot, toward treating research interactions as social interactions per se, is a contribution that interpretivists have made to the measurement literature whose broader implications are still working themselves out.

From a less interpretive perspective, Jennifer Cyr's (2019) work on focus groups offers a parallel demonstration of the expanded range of ethnographic methods in the social science measurement validation literature in recent years. Cyr positions focus groups as a way of evaluating measurement from a qualitative perspective when the concept of interest involves individual-level social interaction; whether about behavior or belief they are thus a way of looking at shared stereotypes, jointly held cultural beliefs, and similar phenomena. In the broader imagination, focus groups are associated with market research; within social science, the most common stereotype views focus groups as a tool for survey and experimental pretests.

Political scientists in particular have underused focus groups as a qualitative tool for investigating group behavior and patterns of agreement/disagreement, which are the distinctive contribution of the method for looking at measurement validation; a single-digit percentage of published articles in the discipline use the method, and around half of those articles focus on individual responses rather than the interactions, shared beliefs, and group dynamics that are the strength of the method in terms of measurement and validation (Cyr 2016). Thus, these approaches still have much more to offer in terms of testing hypotheses and measurement claims about the existence and content of societal consensus, which gives focus groups special relevance in measurement debates regarding culture, ideology, and the interpretation of language and meaning. As such, they remain arguably an underemphasized component of the validation toolkit.

As these illustrative examples show, the set of research tools and perspectives available for immersive qualitative measurement at the individual level in political science is far more developed than it was at the turn of the century – and a complex set of questions about how social structure and hierarchy affect the interactions that produce measurement has been brought squarely into the literature on validity in a way that is productive and worth considering in a broader way. These transformations raise the possibility that the case-based measurement tradition, once usefully regarded as a unitary alternative to multiple quantitative and statistical approaches to validation, might best be seen as divided between a (comparative) historical and an interpretive/ethnographic approach, each offering distinctive methods and raising particular methodological concerns.

HOW MUCH DOES THE SCOPE OF DATA MATTER?

We live in the age of big data, and as a result approaches to social measurement and validation have shifted in interesting, if sometimes surprisingly subtle, ways. In fact, the idea of big data can point to at least two different

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 07:40:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.028

shifts in the landscape of social science measurement. Digital technology has altered the kinds of data that can be created, and it has also shifted the picture of costs and benefits when scholars make choices about measurement validation.

One much-discussed change in the social science landscape over the last decade is the dramatic increase in the amount of available data. Social media, digital measurement, and other aspects of the internet era have made certain kinds of enormous data sets feasible to create and sometimes easier for the research community in general to access. For certain kinds of research tasks, measurement using cell phone data, social media behavior, and other digital interactions can create detailed depictions of real-world dynamics that would have been inaccessible in the recent past.

Alongside these changes that have created new kinds of large-scale data sets, older and more established forms of data have become far more easily accessible. This second change has important consequences for how scholars approach practical issues related to measurement and techniques of validation. Indeed, in some literatures, it has been far more consequential.

Consider, for example, the literatures about measuring democracy that served as the central focus on measurement validation in Chapters 16 and 17 by Collier and co-authors. For the most part, little value has been drawn from social media or massive data sets created with digital devices, largely because these approaches do not capture too much relevant information about whether elections are free and fair, the nature of the franchise, the rule of law, civil–military relations, and so forth. However, practical work related to measuring democracy has seen enormous changes owing in large part to the dissemination of new data sets that compile large numbers of variables in a way that would have been impractical or impossible in earlier eras.

As a central example, consider the impact of the Quality of Government (QoG) data (Teorell et al. 2018) on how scholars approach measurement projects related to democracy. Because the QoG data compile the vast majority of variables that scholars have used to measure democracy as well as related concepts such as participation, corruption, federalism, freedom of the press, and many more, the standard QoG data set includes over 2,000 variables. Until recently, no scholar would have worked with that many variables across an entire career; now, every graduate student who downloads the QoG data for use in a term paper is engaging with a data set whose internal complexity defies human understanding.

This complexity enables, but also requires, subtle shifts in measurement validation practices. Scholars simply cannot have the kinds of close familiarity with the details of cases and measurement process for each of over 2,000 indicators that were once the hallmark of expert use of factor analysis, scaling, and other statistical validation techniques. For example, it was once a common feature of the measurement literature on democracy for scholars to debate the extent to which Freedom House's ties to the US government bias its democracy

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 07:40:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.028

index, with a common hypothesis being that the index is particularly punitive against left-wing governments that are not US allies (e.g., Scoble and Weisberg 1981: 160–61; Nagle 1985: 95; Goldstein 1992: 47; Mainwaring et al. 2001: 145). This concern about the data emerged from knowledge of the staffing and organizational history of the body that produced the data, as well as analysis of the fit between the data and the details of specific cases, and it also motivated a few carefully designed statistical tests (Bollen 1993; Bollen and Paxton 1998, 2000; Steiner 2016).

As the number of available measures per concept has increased, scholars tend to substitute breadth of data for depth of engagement with the fine details of calculation and institutional history behind each indicator. Broader data sets enable the use of statistical measurement models and allow for fairly open-ended processes of discovery, at the expense of the more granular knowledge of the data-generating process that was feasible when fewer variables were available.

For example, Elff and Ziaja (2018) examine twenty-three indicators drawn from four different data sources – a total of 3,863,770 different country-year observations of various subcomponents of democracy scores. It is obviously unreasonable to expect detailed knowledge of each of these measurements. Instead, the authors quite reasonably use a general-purpose statistical model to test for the overall amount of measurement error present in each component of each score, as well as to test hypotheses about whether error is due to information about long-term levels or year-to-year change. They find that the various data sources largely agree on countries' overall level of democracy, particularly with regard to institutional features such as elections and horizontal accountability, but there is much less clarity in the data regarding the timing of change (97–102). This suggests that purely cross-sectional statistical analysis is likely to be on fairly solid measurement ground, but panel models or statistical analyses of regime transition face special measurement challenges.

These findings are very real contributions, providing grounded advice to quantitative users of democracy scores while also identifying key issues for producers of such data. At the same time, it is clear that there has been a shift from depth to breadth in comparison with the earlier literature on potential bias against left-wing governments in Freedom House scores. Elff and Ziaja (2018), along with other similar scholars, demonstrate a great deal of knowledge about democracy measurement – but the pragmatics of an ever-expanding data environment favor generalizable questions and findings that, in the language of Przeworski and Teune (1970: 75–76), substitute variables for proper names. There is nonetheless an undeniable loss of case-study-like textured knowledge in this shift, and it is worth asking whether there can be research designs that combine the excellent breadth of knowledge offered by work such as that of Elff and Ziaja with some of the nuanced local detail of earlier debates.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 07:40:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.028

TEXT MODELS AND MEASUREMENT VALIDATION

While statistical approaches to the study of text have existed in a limited scope for decades, they have become part of mainstream research practice in the social sciences during roughly the last decade. We have seen publications that use text-as-data techniques to study the politics of Muslim clerics (Nielsen 2017), the evolution over time of radical political rhetorics (Karell and Freedman 2019), and the ways that gender shapes politicians' behavior on social media (Beltran et al. 2021). This diversity of applications shows the breadth of impact text-as-data tools have already had on social science measurement and marks them as one of the major changes in this space since 2010. As such, it seems natural to ask whether they have altered the intellectual landscape in terms of measurement validation. I argue that they have, although the changes are surprisingly subtle. Statistical text-as-data tools largely fit within existing measurement validation categories.

While some applications are deeply pragmatic, the majority are straightforward extensions of factor analysis and other techniques from Seawright and Collier's structural equation modeling with latent variables tradition. The data consist of dummy variables representing words (or sometimes combinations of words), and the latent variables or factors can represent different elements for different models, but in social science typically represent themes or ideas. While there are some technical nuances, the conceptual translation is ultimately not too challenging (Grimmer and Stewart 2013).

That said, these techniques do open novel opportunities for mixed-method approaches to measurement validation. Because the data for these methods are in fact text, it is easy and natural for scholars to alternate between quantitative analysis of the data using methods related to factor analysis and qualitative measurement involving in-depth reading of individual documents in relation to the broader corpus.

These mixed-method measurement designs are often undertaken without explicit methodological discussion; scholars simply run a sentiment analysis or a structural topic model, and then select a handful of interesting documents for closer scrutiny as examples of the findings or to make sense of hard-to-interpret patterns. These practices can be reconstructed and expanded in line with the general principle that mixed-method research is best when it explicitly combines methods to compensate for the weaknesses of an initially selected method.

Text-as-data methods have key strengths that are difficult to replicate in qualitative research. They can include huge bodies of text that go far beyond the time and attention constraints of any single human reader. Statistical text methods can also detect subtle patterns in text usage that human readers likely would not see. Finally, they can offer greater levels of consistency across a long project than a single reader can typically sustain – and certainly greater than a team of readers can offer.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 07:40:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.028

At the same time, these methods have important assumptions and limitations. In practice, most applied work with text-as-data models pays little attention to the methods' formal statistical assumptions, which are complex and abstract. Instead, the results are evaluated quite pragmatically; the technique is a success if the results make sense. Scholars evaluate whether the results make sense in two fundamentally different ways. First, they explore whether the quantitative findings are strong and group words together in ways that are theoretically and substantively intuitive. Second, they examine individual documents within and across groupings to determine whether the groupings that classify them together make sense, and to form a qualitative interpretation of the meaning of that grouping. This second approach takes advantage of the strength of qualitative methods in facilitating the creation and interpretation of meaning – and thus stands as a useful example of applied mixed-method measurement.

For example, consider Lacombe's (2021) analysis of how, over the course of the last half of the twentieth century, the National Rifle Association (NRA) taught its supporters to think of themselves as politicized conservative gun activists rather than as a politically diverse collection of people who owned guns. Consistently throughout his book, Lacombe moves back and forth between statistical analysis of texts that the NRA uses to communicate with gun owners, or that those gun owners use to express their political preferences, and qualitative close reading of specific documents as case studies nested within the broader analysis.

In his analysis of the NRA's efforts to construct a Second Amendment ideology of gun freedom, Lacombe begins with a mostly statistical analysis of a structural topic model (89–98), a standard text-as-data model that strongly resembles factor analysis. But to understand the meaning of these topics and their historical evolution over time, Lacombe also offers close readings of carefully selected example texts that represent key usages or innovations. For example, he shows the NRA's early efforts to connect opposition to gun control with anti-communism by quoting a 1935 editorial that imagines Trotskyites saying, "When we have weakened the country by suppressing its rifle bearers … we shall be in a position to go ahead with our plan for setting up a government based on the theories of Karl Marx, Lenin and Stalin" (102). This is not a one-way flow from quantitative to qualitative; Lacombe uses the documents qualitatively to make points about measurement whenever necessary, but readily bounces back to statistical modes as convenient (e.g., 109–12, 115–16, 120–21, 126–27).

Such deeply enmeshed methodological integration is often natural in text-as-data measurement endeavors, because of the nature of the underlying evidence as documents that can in fact be read. At the moment, there is relatively little methodological reflection about such practices, and limited clarity regarding best practices for such mixed-methods measurement. This should be an active research agenda going forward.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 07:40:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.028

CONCLUSIONS

The methods and ideas discussed here fit broadly within the frameworks developed by Adcock and Collier (2001) and Seawright and Collier (2014). The ethnographic and interpretive methods that have taken on fresh importance in recent years fit broadly within the context of the case-based tradition, and scholars working with these methods often (although not exclusively) address questions related to content and nomological/construct validation. Scholars working with large contemporary data sets often use methods drawn from the three quantitative-leaning measurement traditions, and they address tasks from all stages of the measurement process. Finally, as discussed, text-as-data methods largely fit within the structural equation modeling with latent variables tradition and often focus on convergent/discriminant and/or nomological/construct validation. Thus, the frameworks developed in Collier's writings on measurement validation continue to offer a useful map to this domain.

At the same time, recent developments suggest potential refinements in light of technological, intellectual, and social change. There may well be sufficient differences between the ethnographic and interpretive measurement methods that have gained intellectual ground in recent years, in comparison with the more established case-based tradition, to potentially divide the tradition in two. The availability of large, complex data sets through digital distribution pushes scholars toward statistical techniques and, in particular, toward more generalized measurement hypotheses – a trend that leads to good research but also produces some losses. There may be opportunities for research that brings case-based techniques back in to ground these methods in more specificity, although the challenge of how to do this efficiently will demand serious methodological work. Finally, the text-as-data tradition shows scholars who have run ahead of the methodological literature in terms of mixing methods across measurement traditions and the qualitative/quantitative divide. Methodological research that reflects on best practices for this kind of mixture will be an important area in future.

REFERENCES

Adcock, Robert, and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95(3): 529–46. (See Chapter 16, this volume.)

Beltran, Javier, Aina Gallego, Alba Huidobro, Enrique Romero, and Lluis Padro. 2021. "Male and Female Politicians on Twitter: A Machine Learning Approach." *European Journal of Political Research* 60(1): 239–51. https://ejpr.onlinelibrary.wiley.com/doi/abs/10.1111/1475-6765.12392.

Bollen, Kenneth A. 1993. "Liberal Democracy: Validity and Method Factors in Cross-National Measures." *American Journal of Political Science* 37(4): 1207–30.

Bollen, Kenneth A., and Pamela Paxton. 1998. "Detection and Determinants of Bias in Subjective Measures." *American Sociological Review* 63(3): 465–78. www.jstor.org/stable/2657559.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 07:40:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.028

Bollen, Kenneth A., and Pamela Paxton. 2000. "Subjective Measures of Liberal Democracy." *Comparative Political Studies* 33(1): 58–86.
Cyr, Jennifer. 2016. "The Pitfalls and Promise of Focus Groups as a Data Collection Method." *Sociological Methods & Research* 45(2): 231–59.
Cyr, Jennifer. 2019. *Focus Groups for the Social Science Researcher*. Cambridge: Cambridge University Press.
Elff, Martin, and Sebastian Ziaja. 2018. "Method Factors in Democracy Indicators." *Politics and Governance* 6(1): 92–104.
Fujii, Lee Ann. 2017. *Interviewing in Social Science Research: A Relational Approach*. New York: Routledge.
Goldstein, R. J. 1992. "The Limitations of Using Quantitative Data in Studying Human Rights Abuses," pp. 35–61, in T. B. Jabine and R. P. Claude (eds.), *Human Rights and Statistics: Getting the Record Straight*. Philadelphia: University of Pennsylvania Press.
Grimmer, Justin, and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3): 267–97.
Karell, Daniel, and Michael Freedman. 2019. "Rhetorics of Radicalism." *American Sociological Review* 84(4): 726–53.
Lacombe, Matthew J. 2021. *Firepower: How the NRA Turned Gun Owners into a Political Force*. Princeton, NJ: Princeton University Press.
Mainwaring, Scott, Daniel Brinks, and Aníbal Pérez-Liñán. 2001. "Classifying Political Regimes in Latin America, 1945–1999." *Studies in Comparative International Development* 36: 37–65.
Nagle, John D. 1985. *Introduction to Comparative Politics: Political Systems and Performance in Three Worlds*. Chicago: Nelson-Hall.
Nielsen, Richard A. 2017. *Deadly Clerics: Blocked Ambition and the Paths to Jihad*. Cambridge: Cambridge University Press.
Przeworski, Adam, and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. Krieger.
Scoble, H., and L. Wiseberg. 1981. "Problems of Comparative Research in Human Rights," pp. 147–71, in V. Nanda, J. R. Scarritt, and G. Shepherd (eds.), *Global Human Rights: Public Policies, Comparative Measures and NGO Strategies*. Boulder, CO: Westview.
Seawright, Jason, and David Collier. 2014. "Rival Strategies of Validation: Tools for Evaluating Measures of Democracy." *Comparative Political Studies* 47: 111–38. (See Chapter 17, this volume.)
Steiner, Nils D. 2016. "Comparing Freedom House Democracy Scores to Alternative Indices and Testing for Political Bias: Are US Allies Rated as More Democratic by Freedom House?" *Journal of Comparative Policy Analysis: Research and Practice* 18(4): 329–49. https://doi.org/10.1080/13876988.2013.877676.
Teorell, Jan, Stefan Dahlberg, Sören Holmberg, Bo Rothstein, Anna Khomenko, and Richard Svensson. 2018. *The Quality of Government Standard Dataset*. The Quality of Government Institute. Gothenburg, Sweden: University of Gothenburg.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 18 Nov 2026 at 07:40:29, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009551250.028