

Who Rates? National Identity and Democracy Ratings

Sarah Bush* Melina R. Platas†

November 22, 2022

Abstract

Global performance indicators, such as democracy ratings, are influential tools of global governance, and can have a direct bearing on foreign policy, aid, and investment. Many of these indicators rely on expert assessments. Although expert assessments are generally understood to be objective, we suggest that raters' identities may shape their assessments and examine specifically how national identity shapes democracy ratings. Relying on data from two sources, an original survey of experts on Uganda and the Varieties of Democracy Institute, we identify significant differences in ratings provided by national and non-national experts. In most cases, ratings by nationals are significantly more positive, but sometimes they are more negative. We explore three potential reasons why, finding at least some support for each: differences in information access, conceptualizations of democracy, and in-group bias. These findings have implications for our understanding of global performance indicators, which are overwhelmingly a product of Global North organizations.

*Associate Professor (on Term), Department of Political Science, Yale University. Email: sarah.bush@yale.edu. Website: <http://campuspress.yale.edu/sarahbush>.

†Assistant Professor, Social Science Division, NYU Abu Dhabi. Email: mplatas@nyu.edu. Website: <http://www.melinaplatas.com>.

Rankings and ratings of states are influential tools of global governance, not to mention valuable data for research. The growth of these global performance indicators (GPIs) over the past few decades has been remarkable. In the late 1990s there were only twenty GPIs in existence, while by the mid-2010s there were already over eight times as many (Cooley, 2015; Kelley and Simmons, 2019, 493). Much of the literature on this phenomenon focuses on how GPIs affect states, market actors, and inter-governmental organizations (IGOs) and shows that GPIs are an important source of social, political, and economic power.¹

However, these indicators are ultimately created by people, and they are primarily a product of the Global North. Kelley and Simmons (2019, 497) estimate that more than 95 percent of GPIs today are conducted by organizations headquartered in the Global North, particularly Europe and North America. Although there are some signs of change (Swedlund, 2013), Kelley and Simmons' data show that six out of 159 GPIs are produced in Global South countries, while 79 – nearly half – are produced in the United States alone. Thus, although all countries are assessed, actors in very few countries are assessors. The individual raters who participate in the creation of GPIs may come from more diverse backgrounds, but it is almost certainly the case that experts are more likely to be based at Global North institutions.

This phenomenon raises an important question: Do raters' identities affect how countries are evaluated and how ratings are used? Are we coming to systematically different conclusions about country performance as a result of *who* is doing the rating? On the one hand, those who participate in creating GPIs and assessing countries are typically recognized experts and professionals in their respective fields, and methodologies are often specifically designed to limit individual biases. For these reasons, we might expect that rater characteristics will not matter for assessments. On the other hand, there may be conceptual

¹Kelley and Simmons (2015); Dolan (2017); Bisbee et al. (2019); Doshi, Kelley and Simmons (2019); Honig and Weaver (2019); Morse (2019). This literature builds on an earlier sociological literature on ratings and “reactivity.” See Espeland and Michael (2007).

differences even among experts, as well as differences in their sources of information and lived experiences that can inform assessments. As noted by Kelley and Simmons (2019, 505), GPI raters' identities are a lacuna in the literature. Yet they may indeed affect both the content and authority of ratings, as recent research on national perspectives in world politics highlights (Colgan, 2019*b*).

In this paper we hypothesize that rater identity *can* matter for assessments, and examine one potential source of rater subjectivity: national identity. In particular, we explore the relationship between rater nationality and democracy ratings. Democracy ratings are widely used by scholars as well as by non-academics crafting foreign aid policy, non-governmental organization (NGO) advocacy, and investor risk assessments (Girod, Krasner and Stoner-Weiss, 2009; Kaufmann, Kraay and Mastruzzi, 2010, 20; Bradley, 2015, 53). Consequently, they are influential in countries that are the targets of NGO advocacy and recipients of aid and foreign investment (Büthe, 2012; Bush, 2017). Yet despite their scientific presentation, democracy ratings rely on subjective assessments.

To assess whether rater nationality shapes democracy ratings, we rely on two data sources. The first is an original expert survey. The survey focused on democracy assessments of Uganda, a case we chose because – since it is considered a competitive authoritarian regime – it is a most-likely case for expert disagreement about how democratic its regime is. Our survey recruited Ugandan and non-Ugandan governance experts to assess Uganda using a methodology closely modeled on what is used by official democracy ratings. We find that Ugandan experts rated Uganda as significantly freer than non-Ugandan experts. This difference holds when we control for raters' frequency and extent of information consumption about Ugandan politics, perceptions of democratic trends, and demographic characteristics.

The second data source is the Varieties of Democracy (V-Dem) Institute (Coppedge et al., 2011), a leading measure of democracy that relies on a survey of both national and non-national experts. Analyzing the V-Dem data enables us to consider whether the findings

from our original survey generalize to other countries and time periods as well as to an official ratings exercise. Examining the assessments of nearly 2,300 raters across 180 countries and more than 70 V-Dem measures, we find that, more often than not, national raters offered significantly different, and often more favorable assessments, than non-national raters. If anything, however, this pattern is driven by differences in experts' assessments of wealthy, democratic countries, suggesting that it is not only countries similar to Uganda where rater nationality matters. Together with the findings from our original data, the results suggest we should take seriously the question of who participates in GPIs as it shapes the ratings countries receive.

While we are not able to definitively identify the mechanisms underlying the relationship between rater nationality and assessments we propose three explanations, which are not mutually exclusive: differences in information access and exposure; differences in conceptions of democracy; and in-group bias. Our survey of governance experts on Uganda provides some support for each of these mechanisms. For example, our survey documents systematic national differences in experts' views about democracy in the abstract. Nevertheless, the conclusion we draw is not that nationals will always provide more positive assessments of their own countries; indeed, the V-Dem data show that on a substantial number of measures they are more critical than non-nationals. Instead, the key finding is rather that national differences in assessments are possible and even likely. In some cases, co-nationals may be more critical and in others more positive in assessments of their own countries.

This study contributes to our understanding of the reliability and validity of GPIs. Studies have amply documented the social processes that contribute to methodological problems in prominent indicators, including the *Freedom in the World* reports produced by Freedom House (Bradley, 2015; Gunitsky, 2015; Steiner, 2016), the Corruptions Perceptions Index produced by Transparency International (Andersson and Heywood, 2009), and credit ratings such as Fitch, Moody's, and Standard and Poor's (Fuchs and Gehring, 2017).

Why these problems exist has been debated. Some focus on structural issues, such as variable construction and aggregation methods (Bollen and Paxton, 2000; Munck and Verkuilen, 2002). Others, like us, emphasize the characteristics of the individuals or organizations that are responsible for evaluating country performance, including their nationality, political ideology, and cognitive biases and heuristics (Giannone, 2010; Gray and Hicks, 2014; Brooks, Cunha and Mosley, 2015; Colgan, 2019*a*). However, since all these factors plausibly influence the content of GPIs, it has often been impossible to know the counterfactual: that is, what GPIs would look like if different experts were involved or a different approach were used. Our study provides such evidence by first, recruiting a sample of national and non-national country experts to conduct a rating exercise of a particular country and second, examining the association between nationality and democracy assessments using a real-world dataset. Although not all GPIs rely on expert judgments, most do, making our findings of potentially broad relevance.²

In addition, our study speaks to theoretical and normative debates about power and authority in world politics. The conventional wisdom about GPIs is that they gain authority by virtue of their expertise (Kelley and Simmons, 2019, 497). In this way, GPIs' creators are similar to other non-state actors that have managed to gain authority in world politics despite lacking more traditional forms of power (Green, 2013). Yet there is reason to think the picture could be more complicated, as GPIs have sometimes gained considerable influence despite clear limitations in their epistemic quality (Bush, 2017). If ratings' authority in practice depends not only on the competence of their methodologies but also on the identity of raters, then an implication is that theories about authority should pay greater attention to who counts as an expert and not assume that expertise is an objective quality of some

²According to our original coding, 64 percent of the GPIs in Kelley and Simmons (2019) are based entirely or partially on expert judgments. The remainder are produced entirely on the basis of objective, public information (e.g., economic statistics) and are therefore not subject to experts' biases although they may involve other problems (Kerner, Jerven and Beatty, 2017; Dolan, 2017).

raters.

To illustrate some of these dynamics, in a second expert survey we asked a sample of policy elites to assess the authority of different democracy ratings, including those created by our Uganda expert survey. American and European elites reported greater trust in internationally-produced assessments than the assessments by Ugandan experts. By contrast, Ugandan experts trusted our international and Ugandan experts similarly. Thus, who rates matters for the authority of ratings in the real world.

Why Rater Identity Matters

Ratings and rankings are produced by *people*. Yet with a few important exceptions, the literature generally does not theorize or examine the effect of variation in those individuals' characteristics. More attention has been paid to the types of institutions that raters work within; for example, IGOs are noted for being bureaucratic, and this feature influences how they classify countries and fix terms' meanings, as well as their authority when doing so (Barnett and Finnemore, 2004, 31-33).

One explanation for the literature's limited focus on individual raters is that raters' work is expected to be impartial and technical. GPIs' creators are recognized experts in their subject matter, such as democracy, corruption, or financial markets. In some cases, they are trained academics, as with Raymond Gastil, the creator of Freedom House's *Freedom in the World* survey (Bush, 2017, 716), or the World Bank economists who worked on the *Doing Business* report (Besley, 2015, 101-102). In other cases, the experts are non-academic professionals working in IGOs, NGOs, government agencies, and the media. What is common, however, is that expert are members of epistemic communities, or "network[s] of professionals with recognized expertise and competence in a particular domain and an authoritative claim to policy relevant knowledge within that domain or issue-area" (Haas,

1992, 3).

If raters are chosen on the basis of their substantial professional qualifications and training, one might expect variation in their personal characteristics not to cause them to rate countries in systematically different ways. First, one might expect experts to share common ideas about the concepts being rated. After all, shared causal and principled beliefs are closely tied to the development of professional knowledge (Seabrooke and Wigan, 2015). Second, experts might be expected to have access to and rely on similar information about how specific countries are performing. Supporting this logic, Razafindrakoto and Roubaud (2010, 1063-1066) found that experts show some signs of “group think” and ideological bias – but few signs of national or institutional bias – when assessing countries’ levels of corruption. Third, the methods used to produce GPIs may be sufficiently detailed and transparent that there is not opportunity for raters’ individual characteristics to sway their evaluations. Thus, the literature’s implicit starting point is a null hypothesis: *There will be no difference between the country assessments of expert raters with different nationalities.*

We advance a different perspective, arguing that raters’ characteristics – and specifically, their national identity – may affect how they assess countries. Our key hypothesis is thus: *The country assessments of expert raters with different nationalities will differ.*

Our starting point is a recognition that there is often substantial uncertainty, even among experts, about how to evaluate countries on many concepts that are central to GPIs. When it comes to democracy, for example, raters often disagree on how to assess countries that do not easily fall into the categories of consolidated democracy or autocracy. Consider Russia during the 1990s and 2000s. Gunitsky (2015) shows that Freedom House and Polity IV – two leading indicators – came to nearly opposite conclusions about the country’s institutional trajectory and that this sort of discrepancy is common for hybrid regimes in the post-Soviet region. Such differences may be the result of divergent conceptualizations of democracy and measurement strategies. But different raters could interpret the importance, rationale, and

implications of key events or institutions differently even using the same methods due to underlying subjectivity and uncertainty about how to assess democratic performance.

Given that subjectivity and uncertainty, rater identity has the potential to shape ratings. We focus on national identity, although other characteristics (e.g., gender, age) may also matter. Even among experts, nationality can cause individuals to “view the world with particular perspectives and beliefs that shape their perceptions, judgments, and worldviews” (Colgan, 2019*b*, 300) and, more broadly, to use distinct cognitive frames (Cheng and Brettle, 2019). By national identity, we follow other research in referring primarily to the country where an expert was born and lives (Colgan, 2019*b*, 305). In our analysis, we operationalize national identity using measures of both citizenship and residence (which are highly correlated) while acknowledging that the concept is considerably more complex.³

We propose three mechanisms by which rater nationality may shape country assessments: differences in information access, differences in conceptions of democracy, and in-group bias. Although distinct in theory, these mechanisms may be complementary in practice. Below, we elaborate on each mechanism.

First, national and non-national experts may have access and exposure to different information. Especially if they reside in their home country, nationals can read, watch, and listen to a broader array of relevant national media than non-nationals. Although some information sources are available to all experts, content from many smaller newspaper, television, and radio outlets is not available online and may be inaccessible to non-national experts for other reasons (e.g., language). These news sources sometimes cover different issues and events than international media sources, and they may also frame the same events differently. For example, the American media tends to cover other countries in ways that reflect U.S.

³As Colgan (2019*b*, 305) notes, experts’ national identity can be thought of in terms of citizenship, duration of residence in a country, country of education, and other dimensions. It is possible to think about an individual’s national identity (i.e., one’s “Americanness” or “Ugandanness”) on a spectrum (Cheng and Brettle, 2019), which might also be shaped by whether the person belongs to the dominant ethnic or racial group. We leave an exploration of these nuances to future research.

foreign policy priorities and cultural stereotypes (e.g., coverage of Middle Eastern countries often reflects the stereotype that Muslim countries are sexist) (Terman, 2017).

In addition to differences in information access, we expect there to be national differences in information exposure. Nationals tend to follow and consume information about their country more regularly than others. Nationality shapes individuals' social networks and thus the content they are passively exposed to through social media and interpersonal discussion. Given these differences, non-national raters may be more likely to rely on other GPIs as heuristics when forming their assessments.⁴

Supporting this mechanism, previous studies have noted the potential significance of informational dynamics for perceptions of democracy. Logan and Mattes demonstrate that citizens' assessments of countries' democracy are correlated with experts' assessments, but that there are important discrepancies. As they write, "How a political system looks from the ground up can be significantly different from how it looks from the top down" (Logan and Mattes, 2012, 487). It is possible that national experts could share a "ground up" view of democracy in their own countries that is closer to their fellow citizens' perceptions than to non-national experts.

The information mechanism implies national differences in information consumption, knowledge, and country assessments will exist. In terms of the latter, the information mechanism does not make a prediction about the direction of the difference. It is also not obvious how national differences in information will affect assessment quality. One view is that nationals are in a better position to assess their own countries given their better access to local information; this is the perspective taken by V-Dem, which includes mostly national raters (Coppedge et al., 2019, 11). Another is that non-nationals are better able to take a comparative perspective.

⁴Dolan (2017). GPIs encourage this dynamic. For example, Freedom House recommends sources to raters including the Corruptions Perceptions Index, Fragile States Index, and Ease of Doing Business Index.

Second, nationals and non-nationals may evaluate the same country differently because they see the world differently. Many GPIs concern contested ideals, such as development, freedom, and transparency. In our case of democracy ratings, both academics (Coppedge et al., 2011) and practitioners (Hobson and Kurki, 2012) disagree about how to define the concept beyond the minimum of contested elections. Insofar as a rating system provides detailed rules about how to evaluate countries, these disagreements matter less, but the challenge of eliminating all subjectivity in expert ratings is significant.

Thus, it is plausible that individuals' lived experiences will prompt them to interpret the concepts being rated in ways that reflect their education and experiences. For example, a situation or incident that seems to be a clear indicator of democratic backsliding to one person may seem less problematic or as a continuation of the status quo to someone else, perhaps because they have personally experienced a much worse situation or view the most important components of democracy differently. We know from surveys of citizens in nineteen African countries that people who have experienced a leadership turnover in their country have higher standards for democracy when asked to evaluate the quality of hypothetical elections (Bratton, 2010).

Experts may also evaluate countries differently because they have different values. For example, nationals of some countries may be more likely to understand the concept of "democracy" to include socioeconomic equality, whereas others may think about it in terms of collective security (Schaffer, 2000). Research suggests that political orientation develops early and can be sticky (Sears and Brown, 2003). For example, immigrants' ideas about governance may vary systematically with the type of political regime in their country of origin (Bilodeau, 2014; Just, 2019). This literature suggests that variation in experts' nationalities will influence their assessments of country performance; we draw out observable implications about the distinction between nationals and non-nationals in particular given our focus in this paper.

The second mechanism implies that nationals and non-national experts will evaluate the same country differently, but similar to the first mechanism, it does not predict the direction of that difference. Another observable implication is that there will be national differences in terms of how experts conceptualize democracy in the abstract, beyond their assessments of a particular country’s democratic trajectory.

Finally, nationals and non-nationals may evaluate the same country differently because of the phenomenon of in-group bias. Whereas the first two mechanisms simply imply that nationals and non-nationals will conceptualize and thus evaluate democracy *differently*, this mechanism implies that nationals will provide *more positive* assessments of their own country than non-nationals. The same logic implies that non-nationals may provide more critical assessments of other countries as an expression of negativity towards an out-group. Thus, to be clear, we do not take a position about which group of raters is more accurate.

Drawing on research by Colgan (2019a), this mechanism rests on the well-known finding from social psychology that people display in-group favoritism. One explanation for this dynamic is through the individual’s desire for self-esteem; people want to view themselves, and therefore the groups they belong to, positively. The national in-group is an especially salient political identity (Mutz and Kim, 2017).⁵

Due to in-group favoritism, we might expect even experts to evaluate their own country more favorably than out-group raters. Since democracy is often perceived as a normatively-desirable trait, the logic of in-group favoritism implies that in-group experts will view their own country as more democratic than out-group experts. Moreover, it implies that nationals will assess their own country more favorably *in general*. Whereas the other two mechanisms explaining why national identity could affect ratings do not necessarily predict that nationals and non-nationals will reach different assessments across the board (e.g., on

⁵In-group favoritism can extend to other countries. Studies have found that experts view their countries’ allies as more democratic than is merited (Oren, 1995; Colgan, 2019a).

all sub-components of democracy), the in-group favoritism mechanism expects more consistent differences. Although such a bias is surprising if we consider expert raters to form an epistemic community with a high degree of internal cohesion and professionalism, this type of behavior is more plausible if we recognize that expert raters are less cohesive as a group, perhaps because they have been trained at different institutions, meet infrequently, and lack a common culture (Cross, 2013, 150).

Again, there is evidence that supports this mechanism from the level of citizens.⁶ Large majorities in non-democratic countries in East Asia such as Singapore and Vietnam perceive their countries as democratic (Pietsch, 2015). Similarly, more than three-quarters of respondents surveyed in Rwanda, Cambodia, and Thailand say the media in their countries have “a lot of freedom,” but all three countries have been coded by Freedom House as “not free” and are known to have substantial restrictions on the press (English and Becker, 2013). Although these patterns are also consistent with social desirability bias or fear of criticizing the government, they suggest a tendency to believe one’s own country is democratic.

Raters themselves are attuned to this possibility of in-group bias. For example, international election observation missions generally do not include monitors from the countries that are holding elections (Carothers, 1997, 27; Hyde, 2012, 50). The logic is that by excluding country nationals, international observers’ reports will be (or will be perceived as) more accurate.

Research Design

To test our hypothesis against the null, we use two types of data. First, we examine democracy assessments from an original survey of national and non-national experts on Ugandan politics. Second, we analyze V-Dem experts’ ratings, comparing the country evaluations of

⁶At the same time, we note that public attitudes related to democracy are strongly correlated with expert surveys according to several analyses (e.g., Bratton, Mattes and Gyimah-Boadi, 2005; Norris, 2013).

national and non-national experts across all countries over a range of years.

Each data source has advantages. Since we conducted the Uganda expert survey ourselves, we could include questions designed to examine the plausibility of the three mechanisms, explore experts' views about democracy in the abstract, and address potential confounders. This survey also allows us to look in-depth at how experts evaluate a competitive authoritarian regime's qualities. Hybrid regimes are considered most-likely cases for systematic disagreements in ratings due to the ambiguity of how to classify them. They are also the most common regime type among non-democracies in the world today and have become much more common in recent years (Lührmann, Tannenberg and Lindberg, 2018, 8).

Our second source is data from the Varieties of Democracy Institute, which allows us to examine the relationship between rater nationality and democracy assessments over time and across a variety of indicators and countries. This analysis addresses two concerns about the external validity of the Uganda expert survey.⁷ The first is whether our findings generalize to other countries and time periods – and to a larger sample of experts. The second is whether the findings would be similar in a real-world rating, although we attempted to conduct our expert survey in such a way that approximated one. That our findings about national differences from the V-Dem data largely corroborate our findings from the Uganda expert survey provides further support that nationality may indeed matter even for expert assessments.

Uganda Expert Survey

To examine whether raters' national identities affect their democracy assessments, we recruited a sample of Ugandan and non-Ugandan experts on Ugandan politics to participate in an online survey in 2019. Our analysis of their assessments follows the plans we pre-

⁷Ideally, we would have conducted a similar analysis using official Freedom House data. Doing so was not possible because the organization does not make information about its coders available.

registered with the Evidence in Governance and Politics repository unless otherwise noted.⁸ First we provide our logic of case selection and background information.

Case Selection and Background

Uganda is considered a competitive authoritarian regime. Its aggregate Freedom House score – on which a rating of “not free,” “partly free,” or “free” is possible – is around the cutoff between “not free” and “partly free.” Its score is also near key cutoffs on Polity IV, which assigns countries a score between -10 (hereditary monarchy) and 10 (consolidated democracy). Uganda is rated -1, at the border between the Polity categories “closed anocracy” and “open anocracy.”

Uganda has been governed by the the National Resistance Movement (NRM) and President Yoweri Museveni for nearly 40 years. Prior to that, Uganda experienced periods of political instability and civil war between independence in 1962 and 1986, when the National Resistance Army overthrew the previous government. There has been no peaceful transfer of power at the executive level since independence.

Yet since the NRM took power, there have been some institutional changes that could be interpreted as democratizing. For example, after nearly two decades of “no party” rule, a 2005 national referendum reinstated multiparty elections. Four multiparty elections have been held since. The elections are competitive, as a large percentage of incumbents lose their seats at the parliamentary and local government levels. At the same time, however, the parliament voted to eliminate presidential term limits in 2005 and eliminate an age limit on the presidency in 2018. Together, these changes have allowed Museveni to continue participating in elections. Presidential elections have some competition, with opposition candidates garnering up to 40 percent of the vote, but Museveni has continued to win.

⁸The analysis in the paper follows the pre-analysis plan unless otherwise specified; see (SI) §1 . See SI §8 for a discussion of research ethics related to our data collection.

Further, the NRM has dominated parliament since the reinstatement of multipartyism. In sum, Uganda holds elections with some competition, but the playing field is tilted in favor of the president and his party in key ways.

Despite the fact that democracy ratings regularly rate Uganda as less than a full democracy, most citizens think otherwise. Figure 1a shows the percent of Ugandan respondents across seven Afrobarometer survey rounds who assess Uganda as a democracy, as well as those who are uncertain.⁹ The vast majority of Ugandans surveyed view Uganda as a democracy, and these assessments have been relatively stable. By contrast, Figure 1b shows that Uganda has not received a Polity IV score that would categorize it as a democracy (6 or above) since independence, and Freedom House has never categorized Uganda as “free.” The discordance between citizens’ and experts’ assessments is consistent with the idea that nationals hold distinct (and potentially more favorable) perceptions about democracy in their country, although we acknowledge the potential for social desirability biases in the survey responses.

Uganda is also precisely the type of country for which democracy ratings matter. It has relied on foreign aid for as much as 70 percent of government expenditures since the 1990s (Findley et al., 2017, 642). As such, perceptions about democracy inform donors’ decisions about aid and other “democracy-contingent benefits” (Hyde, 2011, Ch. 3). In December 2020, for example, the chairman of the U.S. House of Representatives Foreign Affairs committee wrote a letter to the Secretaries of State and Treasury citing “the alarming slide towards authoritarianism in Uganda” as a rationale for reviewing all non-humanitarian aid (Engel, 2020). Thus, expert perceptions of democracy in Uganda have the potential to affect domestic stability through their effects on aid, sanctions, and other policies.

⁹SI §2 includes a figure with an alternative coding of responses.

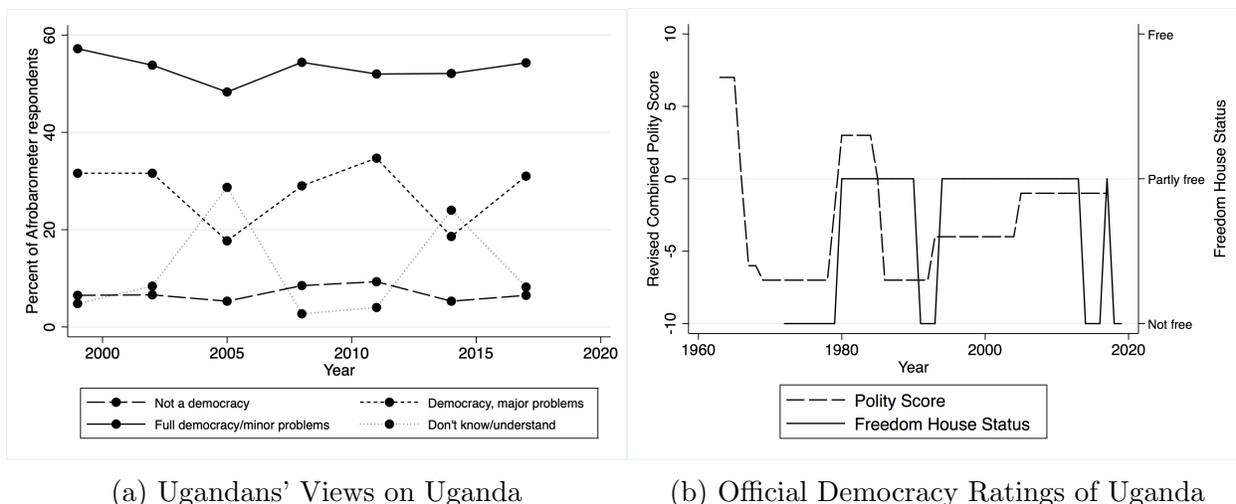


Figure 1: **Citizen and Expert Assessments of Democracy in Uganda.** Data are from Afrobarometer Rounds 1-7, Polity IV, and Freedom House. In the left panel, we combine into “Full democracy/minor” two answer categories about the extent of democracy: “full democracy” and “A democracy, but with minor problems.”

Expert Sample

The population of interest for our survey is people who have sufficient expertise in Ugandan politics that they could serve as raters for a democracy GPI such as Freedom House. Our interviews with people involved with the creation of democracy GPIs¹⁰ and a review of their publicly-available information indicate that the experts who evaluate countries for Freedom House and other GPIs come from both academia and other fields, where they are often program officers at NGOs engaged in democracy promotion or fellows at think tanks (e.g., the United States Institute of Peace).¹¹ As such, we targeted individuals with similar backgrounds. In total, 220 experts were invited via e-mail to take the survey, of whom 118 answered at least some questions for a 54 percent response rate. Though our sample is necessarily small, its size and the survey response rate both compare favorably with related

¹⁰We interviewed five people involved in the creation or production of prominent democracy GPIs. These interviews were conducted on off-the-record terms. Further details are discussed in SI §8.

¹¹For the list of raters for *Freedom in the World 2020*, see the “analysts” acknowledged in the report: <https://freedomhouse.org/fiw-2020-acknowledgements> (accessed July 31, 2020). On how potential V-Dem raters are chosen, see Coppedge et al. (2019, 58-61).

expert surveys.¹²

To identify non-Ugandan experts, we used several methods. First, we reviewed the websites of top American and European academic departments in political science and African studies, using “Uganda” as a search keyword. Additionally, we identified experts from the United States, Canada, and Europe who have authored or co-authored works on Uganda by consulting past programs from the American Political Science Association meetings and using Google Scholar searches. Individuals entered the sample if they were academics who had published multiple works on Uganda in the past ten years or if they were policy professionals whose work portfolios involved Uganda. Based on interviews with people involved with prominent democracy GPIs, we believe this process closely resembles their recruitment processes, with the exception that we could not mimic their social network dynamics over time (e.g., when last year’s expert suggests a colleague to replace her this year).

To identify Ugandan experts, we focused on individuals working in several sectors who are knowledgeable about Ugandan politics. First, we included lecturers and faculty members at the political science department (or equivalent) across four Ugandan universities. Second, we included the senior staff of major governance-related NGOs and other important policy analysts based in Uganda. We included editors and senior journalists at the major media houses, most of which are independent, as well as political commentators at prominent news outlets. Although journalists are likely not included as often as academics, they are highly knowledgeable about governance and politics and are, by profession, trained to be balanced.

The people in our sample reported considerable knowledge about Uganda, as expected. 70 percent reported following Ugandan politics every day, and a further 22 percent reported following it a few times a week. 87 percent said that they were “somewhat” or “very” certain in their responses to questions about democracy in Uganda. We exclude respondents from

¹²For related examples, see Gray and Slapin (2012, 317), Dietrich (2016, 90), and Swedlund (2017, 466). Additional notes on sampling are in SI §2.

the analysis if their responses indicated they were not in fact experts.¹³

The samples of Ugandan and non-Ugandan experts differed in ways beyond their nationalities as explored in SI §3.1. The non-Ugandan experts who participated in our study were more likely than Ugandan experts who participated to be women, work in academia, and possess doctoral degrees. They were also farther left on a left–right ideological spectrum. Although it is not obvious to us whether or how these background characteristics would be expected to influence raters’ assessments of democracy in Uganda, we control for them in our analysis as potential confounders of the relationship between nationality and ratings.

Methods for Assessing Democracy

Our expert survey was completed anonymously using the online platform Qualtrics. We doubt that many respondents would have felt concerned about government reprisals or social desirability in what was (truthfully) described in the recruitment materials and informed consent as an academic study by the authors about democracy in Uganda. Nevertheless, the survey mode (anonymous and online) should have minimized such concerns.

To have the experts assess democracy in Uganda, we used a similar methodology to the one developed by Freedom House. *Freedom in the World* assesses states’ levels of freedom and provides both numerical scores for political rights and civil liberties using a 100-point and also overall ratings of countries as “free,” “partly free,” and “not free.” They are widely used, not only by scholars but also by government officials, journalists, and firms, especially in the United States (Bush, 2017). As such, they merit exploration in their own right. However, the Freedom House case can also inform our understanding of the many other expert surveys of democracy (including Polity, V-Dem, and the Economist Intelligence

¹³Our pre-analysis plan specified we would exclude respondents who “never” or “rarely” read about Ugandan politics or felt “very uncertain” answering questions about democracy in Uganda. Only one person “rarely” read about Ugandan politics (and none “never” did), and only eight felt “very uncertain.” After fielding the study, we decided to further limit the sample to respondents who knew the name of the speaker of the Ugandan parliament, which further eliminated thirty-one respondents (26 percent).

Unit) as well as other GPIs that rely on this method (such as the Corruption Perceptions Index and Ease of Doing Business index).

The Freedom House questionnaire currently involves twenty-five questions: ten questions about “political rights,” and fifteen questions about “civil liberties.” We used the same questions, asking raters in 2019 to assess Uganda in 2018; our survey took place in early 2019, prior to the release of the 2018 *Freedom in the World*. Each question involved giving a country a score between zero (least freedom) and four (most freedom) such that the maximum possible score for a country was 100. Using the published methods of Freedom House, we then computed overall scores for political rights and civil liberties based on respondents’ answers. The questions cover issues such as whether elections are free and fair and the media is independent are listed in SI §4.

Although we made every effort to follow the methodology used by Freedom House to promote comparability, two differences arose. First, the stakes of the rating exercise were lower. If raters were assessing countries more quickly in our study, they may have been more influenced by their national identity. To encourage participants to take our survey seriously, we appealed to their professionalism and offered a monetary incentive worth US\$20. In fact, we believe the experience may have been fairly similar to several prominent GPIs. For example, V-Dem asks country experts to take an online survey for US\$25 compensation.¹⁴

Second, individual raters’ decisions were not debated by others.¹⁵ Freedom House accomplishes this task through ratings review meetings at the regional level. We may overestimate national differences by skipping this review step, although insofar as reviewers at the regional level are less likely to be nationals of the country in question than the raters themselves, the

¹⁴See <https://www.v-dem.net/en/news/call-country-experts-v-dem-update-2020/> (accessed July 31, 2020).

¹⁵For a further discussion of how the Freedom House methodology attempts to prevent unwarranted year-to-year changes in countries’ scores and how our survey explored this theme, please see the discussion in our pre-analysis plan (SI §1).

review process could exacerbate such differences.

Main Findings

To test whether ratings differ by rater national identity, for the main analysis we classify respondents as Ugandan or non-Ugandan based on their reported citizenship. In SI §3.2, we code the national in-group on the basis of residency. In practice, there were few Ugandans in the sample who were not residents of Uganda, and vice versa.

We evaluate the effect of nationality on democracy assessments using answers to the twenty-five questions used by Freedom House. We summed together the responses following the reported Freedom House procedure to create a numerical assessment on which 100 is the maximum possible (i.e., most democratic) score. We find significant differences in ratings between national and non-national experts, with Ugandan experts providing higher ratings on average ($p = 0.009$). Figure 2 shows the distribution of scores across the two sets of raters, with Ugandan raters shown in blue and non-Ugandan raters shown in grey, and the mean scores for each group indicated by vertical lines. This pattern supports our hypothesis that national and non-national ratings will differ. Although national raters were significantly more positive in Figure 2 than non-national raters, they were still fairly critical.

This difference is large. National experts assessed Uganda about eleven points higher than non-national experts on the 100-point scale, which is equivalent to a 31 percent increase in the overall score. In fact, the score provided by non-nationals was on the border of “not free” and “partly free” according to the Freedom House methodology, whereas the nationals’ score was comfortably within the “partly free” category. The numerical scores are particularly important since some of the most influential users of democracy GPIs focus on them. For example, the Millennium Challenge Corporation, a U.S. foreign assistance program, uses twenty-five as the cut-off for scoring countries as passing in terms of civil liberties; non-Ugandan experts in our survey barely thought the country passed with an average score of

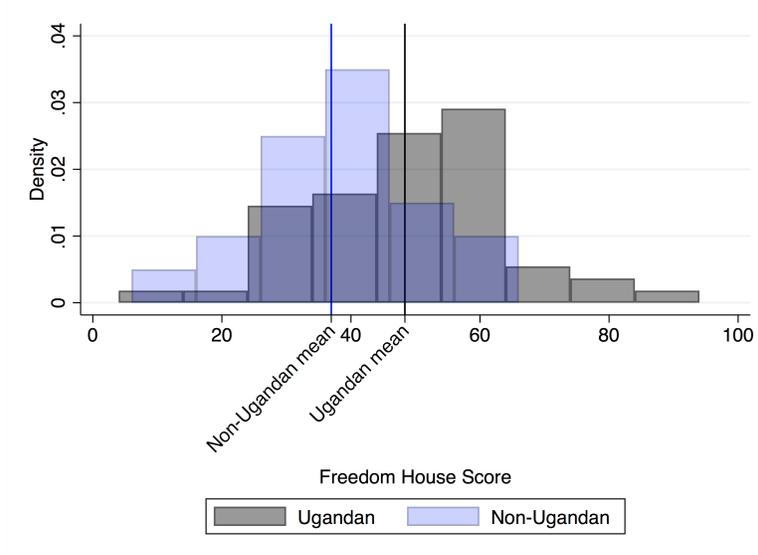


Figure 2: **The Distribution of Democracy Scores by Rater Nationality.** This figure shows experts’ raw scores, with a score of 100 indicating greatest democracy. Difference = 10.51, $p = 0.006$ according to a two-tailed t-test with equal variances.

twenty-six, whereas Ugandan experts’ average was thirty-four.¹⁶

As noted earlier, Freedom House groups its questions into two categories: political rights and civil liberties. We next examine national differences in both sets of questions, as well as the overall score, using ordinary least squares (OLS) regressions.¹⁷ For this analysis, we rescaled all outcome measures to be between zero and one. We also introduce measures for variables that could plausibly confound the relationship between national identity and democracy ratings, including those related to other rater demographic characteristics (gender, political ideology, and profession) and knowledge about Uganda (frequency of following Ugandan politics, and certainty in responses, and number of Ugandan media sources followed).¹⁸ Nationality is undoubtedly a complex “treatment” that likely involves additional differences beyond what we can or did measure; this complexity is part of why we find the

¹⁶See <https://www.mcc.gov/who-we-select/indicator/civil-liberties-indicator> (accessed March 3, 2022).

¹⁷SI §3.4 contains the full regression results.

¹⁸We show in SI §3.5 that the results are similar, though more marginal in terms of statistical significance, when we include only demographic control variables.

topic worthy of study. Nevertheless by controlling for these variables, we are able to hold constant some of the most plausible confounders. As shown in Figure 3, we find that Ugandan raters provided higher scores for both the political rights and civil liberties sections of the Freedom House questionnaire, but the difference was only statistically significant at conventional levels for the civil liberties questions when controlling for relevant covariates.

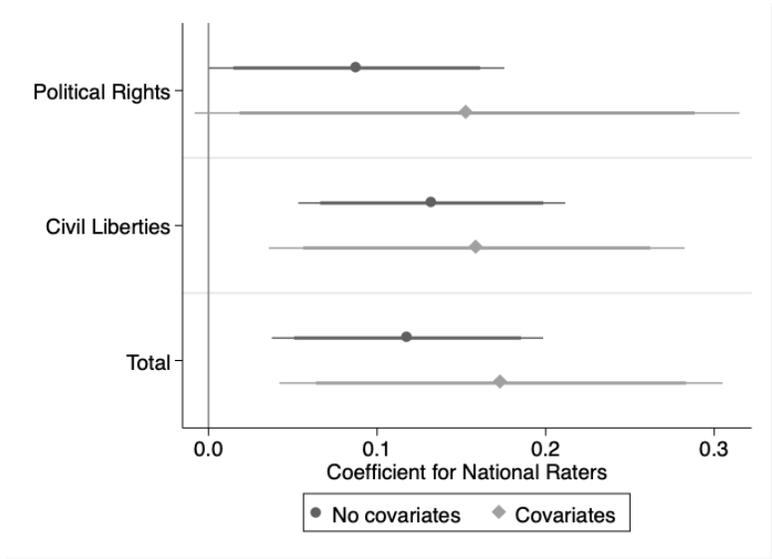


Figure 3: **The Relationship between Rater Nationality and Democracy Assessments, by Sub-indicator.** The figure displays coefficient estimates (with 95 and 90 percent confidence intervals) based on OLS regressions of democracy assessments on an indicator for the respondent’s national identity. The covariates included in this analysis are rater gender, left–right political ideology, frequency of following Ugandan politics, profession, certainty in responses, and number of Ugandan media sources followed.

As a robustness check, in SI §3.3 we re-run the analysis after removing respondents with scores which could be considered outliers, i.e., in the bottom 5 percent and top 95 percent of the distribution of raw scores. Our results are robust to this specification and in fact slightly more statistically significant.

V-Dem Democracy Assessments

To what extent do our findings from Uganda experts travel to other countries and times? Would we observe these differences in a real-world as opposed to artificial assessment scenario? To address these concerns, we turn to data from the V-Dem. For most democracy GPIs, including Freedom House, information about coders' nationalities is either not maintained or not accessible to researchers. By contrast, V-Dem gathers such information about its thousands of raters via a post-survey questionnaire. For this reason, and also because of the project's prominence and quality, the V-Dem assessments provide an excellent additional data source. Our analysis builds on that of Colgan (2019*b*, 364-365), who compared American and non-American coders' assessments of vote buying in countries that are friendly, hostile, and neutral towards the United States, finding national differences.

Description of the Nationality Data

V-Dem is a massive coding enterprise in which expert raters are invited to evaluate countries according to hundreds of questions. The raters may code one country over time, multiple countries over time, or multiple countries at one point in time. V-Dem aggregates the individual coders' ratings into both specific indicators (e.g., for free and fair elections) and higher-level indices (e.g., for liberal democracy) using Bayesian methods.

Crucially for our purposes, V-Dem raters are asked to fill out a post-survey questionnaire which covers, among others, whether raters were born in the main country they rated and whether they currently reside there. The V-Dem data for which we have information on coder nationality covers nearly 2,300 raters across 180 countries.¹⁹ Out of consideration for raters' privacy, V-Dem does not provide information on where the rater was born or resides if not

¹⁹Notably, we are not able to include those raters who evaluate more than one country as we cannot determine their nationality from the data we have access to. Unfortunately there is no way for us to determine how similar or different they are from raters of only one country and thus no way to assess the extent of bias due to their exclusion.

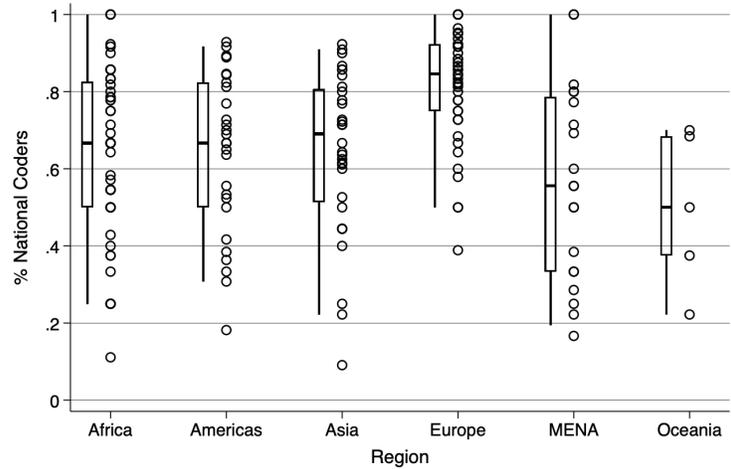


Figure 4: **Variation in Reliance on National Raters by Region.** Rater nationality is coded based on citizenship from the V-Dem post-survey questionnaire.

in the rated country. Although the post-survey questionnaire also gathers other information about raters' identities (e.g., age, gender, and education, among other variables of interest), we could not access this information. As such, we cannot assess the extent to which national or resident raters for a given country–year differ from non-national or non-resident raters in terms of other characteristics.

As Figure 4 shows, there are intriguing regional differences in V-Dem's reliance on national raters. Although most V-Dem experts are nationals in every world region, European countries are more likely to be coded by nationals, and countries in the Middle East and North Africa (MENA) and Oceania regions are less likely to be coded by them. This pattern may reflect regional variation in the availability of individuals with the requisite credentials, the integration of such individuals into the scholarly networks used for recruitment, or other factors (Coppedge et al., 2019, 59).

Main Findings

We analyze two types of outcome variables from V-Dem. Both suggest that there are significant national differences in democracy assessments. First, as part of the V-Dem post-survey questionnaire, all raters are asked to provide an assessment of democracy in the main country they rated as of 2012 (*v2zzdemsc*). The instructions from V-Dem begin: “Imagine a scale that measures the degree of democracy–autocracy in countries around the world, stretching from 0 to 100. 0 represents the most extreme autocracy in the world and 100 represents the most democratic country in the world.” We refer to this as the “democracy score”. V-Dem does not ask raters as part of its main survey to provide an overall democracy assessment by design, since it uses a Bayesian measurement model that aggregates across respondents and indicators to create indices. The responses to the democracy score question from the post-survey questionnaire enable a relatively comparable analytical approach to what we used in the Uganda expert survey and to what is used by some other expert-created GPIs.

We present the results from OLS regressions of the raters’ democracy scores on their nationality and residence in Table 1. National raters assessed their main country of expertise to be significantly more democratic than non-national raters; the difference is about eight points on the 100-point scale in the first model, a nearly 14 percent increase. The difference is even larger when considering residents versus non-residents – a fourteen point or 25 percent increase in the democracy score. When we introduce regional fixed effects in the second model, mindful of the patterns revealed in Figure 4, we see that there is still a significant national difference, although the size of the difference is more than cut in half. The inclusion of country fixed effects reduces the coefficients and statistical significance further, though there is still a significant difference when comparing residents and non-residents. The fact that the relationship is weaker when including region and country fixed effects likely reflects how the nationality effect is driven by particular countries and regions, a point to which we return below.

	(1) National	(2) Resident	(3) National	(4) Resident	(5) National	(6) Resident
Coder is a national	7.900*** (1.294)		3.452*** (1.220)		1.059 (1.062)	
Coder is a resident		13.757*** (1.185)		8.338*** (1.118)		2.063** (0.955)
Constant	57.932*** (1.117)	55.125*** (0.994)	51.356*** (1.443)	49.006*** (1.370)	63.412*** (5.324)	62.656*** (5.306)
Region FE	No	No	Yes	Yes	No	No
Country FE	No	No	No	No	Yes	Yes
N	2,241	2,238	2,232	2,229	2,241	2,238
R ²	0.018	0.062	0.219	0.235	0.604	0.605

Table 1: **The Relationship between Democracy Scores and Rater Nationality and Residence.** Democracy scores refer to the rater’s assessment of democracy in their main rated country for the year 2012 on a 0-100 scale. Robust standard errors are in parentheses.

Second, as part of the main V-Dem questionnaire, raters were asked to assess many different potential indicators of democracy. We focus on the indicators that contribute to at least one of five V-Dem democracy indices (electoral, liberal, participatory, deliberative, or egalitarian). The relevant indicators for our analysis are the ones that involve subjective assessments (e.g., “How restrictive are the barriers to forming a party?”), which V-Dem aggregates using Bayesian methods, not those that involve factual information (e.g., “What is the minimum age at which citizens are allowed to vote in national elections?”). There are 71 such indicators. We further restrict this analysis to coders’ main countries, which has the advantage of partially controlling for the coders’ levels of expertise since we focus on their most-expert country. Importantly, our analysis can now include fixed effects for the country *and* year. Such an approach is possible because the V-Dem project involves multiple coders’ assessments for each indicator for a given country–year.

Table 2 summarizes the results of our analysis of the 71 official V-Dem indicators; SI §5 contains the full results. There were significant national differences on more than three-quarters of the indicators. Most commonly, national raters were more favorable, similar to

Category	Number (%) of sub-indicators where nationals more negative	Number (%) of sub-indicators where nationals more positive	Number (%) of sub-indicators with no difference
Elections	0/12 (0%)	9/12 (75%)	3/12 (25%)
Political Parties	0/4 (0%)	2/4 (50%)	2/4 (50%)
The Executive	1/5 (20%)	1/5 (20%)	3/5 (60%)
The Legislature	0/4 (0%)	4/4 (100%)	0/4 (0%)
Deliberation	3/7 (43%)	4/7 (57%)	0/0 (0%)
The Judiciary	1/4 (25%)	2/4 (50%)	1/4 (25%)
Civil Liberty	5/19 (26%)	9/19 (47%)	5/19 (26%)
Civil Society	2/5 (40%)	2/5 (40%)	1/5 (20%)
The Media	3/6 (50%)	1/6 (17%)	2/6 (33%)
Political Equality	1/5 (20%)	4/5 (80%)	0/5 (0%)
Total	16/71 (23%)	38/71 (54%)	17/71 (24%)

Table 2: **The Relationship between V-Dem Indicators and Rater Nationality.** All regressions include country and year fixed effects and robust standard errors. Difference is coded using a threshold of $p < 0.05$.

the finding about the direction of difference in both the democracy scores and the Uganda expert survey. But in many cases, they were more critical. The overall picture that these patterns paint is one in which national differences among expert raters do matter. However, the fact that differences do not show up consistently across all indicators and they are not always in the direction of national favoritism cuts somewhat against in-group bias as a potential explanation.

Where are national differences are most commonly observed in the V-Dem data? Returning to the democracy scores outcome measure, Figure 5 shows how the relationship between experts' nationality and assessments varies as countries become wealthier. This figure is based on an OLS regression of the score on the coder's nationality, the country's GDP per capita, and their interaction. Although national raters perceive countries to be significantly more democratic than non-national raters when countries are wealthy, the relationship reverses when countries are poor. As would be expected given the relationship between economic development and democracy, we show in SI §5 a similar interaction be-

tween nationality and country Polity score.

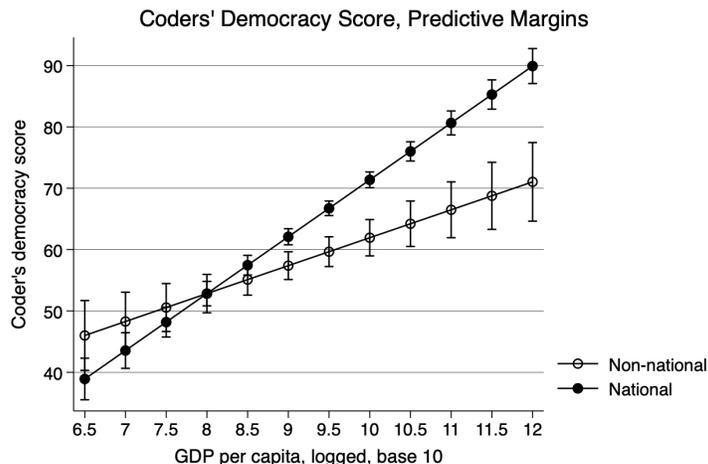


Figure 5: **The Marginal Effect of Country GDP and Coder Nationality on Democracy Assessments.** Democracy scores refer to the rater’s assessment of democracy in their main rated country for the year 2012 on a 0-100 scale. The results are based on an OLS regression of the assessment on the coder’s nationality, the rated country’s GDP per capita, and their interaction with robust standard errors.

Unfortunately, limitations in the V-Dem data prevent us from making inferences about the drivers of variation in national differences across different types of countries. For example, the answer may lie in compositional differences in the types of people who tend to be raters for countries in different regions. We can draw an important conclusion from this analysis, however: it does not seem to be the case that non-national raters are especially critical of lower-income countries or African countries. Thus, we do not suggest that our findings from the Uganda expert survey necessarily indicate a broader tendency for national raters in Uganda or other African countries to display national favoritism. If anything, Figure 5 suggests that Uganda could be atypical of the types of countries where positive national bias is most likely to be observed globally, although we acknowledge that this conclusion is only based on an analysis of one GPI.

Exploring the Mechanisms

Thus far, we have established that national and non-national experts rate Ugandan democracy differently, and that these differences are also observed for a wide variety of indicators in the V-Dem data, contrary to the null hypothesis that national identity is not relevant. The next question is why. Unfortunately, given the absence of further information about who V-Dem raters are, that data source cannot shed further light on the plausibility of the information access or conceptual differences mechanisms we theorized earlier. However we are able to offer three further analyses using the Uganda expert survey to shed light potential mechanisms.

First, to consider the information mechanism, we draw on questions about the experts' consumption of information about Ugandan politics. We posited that national experts may consume more and different amounts of political information than non-national experts. As expected, non-Ugandan nationals follow fewer Ugandan news sources ($p < 0.001$), know less about Ugandan politics according to our political knowledge questions ($p < 0.001$), and are less likely to be certain about their responses to questions about democracy in Uganda ($p < 0.001$). These patterns support the idea that significant informational differences exist and could explain experts' different ratings. At the same time, we note that national differences persist in Figure 3 after controlling for those variables. Thus, information access may be an incomplete explanation of national differences, although our measures may not capture the full extent of variation in information access.

Second, we consider on which questions national and non-national experts provided significantly different responses. SI §3.6 lists the full text of the questions. Nationals provided higher (i.e., freer) assessments on all but one of the questions for which there are significant differences across the two groups. Most questions where we identify differences are in the civil liberties section of the questionnaire, especially on questions relating to personal auton-

omy and individual rights. Notably, there are no differences on assessments for the questions about the electoral process or functioning of government. There is also one dimension on which non-Ugandan raters assessed the country as freer: the extent and effects of outside influence on Ugandan politics. Insofar as the in-group bias mechanism predicts that national raters will give their country higher ratings across the board, these patterns are potentially more consistent with the other mechanisms. At the same time, we do not see much difference in the magnitude of the coefficient for nationality across the different categories – political rights and civil liberties – in Figure 3.

Finally, we consider responses to questions about democracy in the abstract. Using the anchoring vignettes developed by King et al. (2004), we asked the experts to consider six hypothetical scenarios concerning two important dimensions of democracy: competitive elections, which are a core dimension of democracy even in its most minimalist definitions (Coppedge et al., 2011, 254); and freedom of speech, which is associated with more maximalist definitions of liberal democracy, including the definition used by Freedom House.²⁰ For each scenario, respondents indicated how free they thought the country was, with the response options ranging from zero (not free at all) to four (completely free).

National and non-national raters perceived the hypothetical scenarios in notably different ways. Table 3 shows the scenarios and national differences for the competitive election scenarios; a similar table for freedom of speech scenarios can be found in SI §3.7. Positive differences indicate that the non-national raters evaluated the scenario as being freer than the national raters.

For example, while the non-Ugandan experts we surveyed had a fairly strong consensus that the first scenario was between “very free” and “completely free” (with an average score of 3.3), Ugandan experts were somewhat more equivocal, while still judging the scenario to

²⁰The vignettes were randomized such that they appeared either before or after the questions concerning democracy in Uganda.

Scenario	Difference
James does not likely many of the government’s policies. An election is coming up for the legislature at the regularly scheduled time. He is planning to vote for a party that has been campaigning in the media and in his town to change the governments policies.	0.83, $p = 0.002$
Matthew disagrees with many of the government’s policies. Although it is difficult for his favored political party to get covered in the media, it will be running a full slate of candidates in the upcoming legislative election. He is still planning to vote for this party.	0.62, $p = 0.009$
Ronald has political views at odds with the government. The government has called for a legislative election on an accelerated schedule that has made it difficult for opposition candidates to campaign. He was able to register to vote and is still planning to vote for an opposition party.	0.74, $p = 0.002$
Andrew does not like the government’s stance on many issues. He is planning to vote in the upcoming legislative election, although the party he likes most has had several of its candidates arrested on false charges.	0.28, $p = 0.078$
Charles disagrees with many of the government’s policies. There is going to be a legislative election soon, but international election monitors have not been invited and domestic independent election monitors have been jailed. He does not believe that the official results of the vote will be reported honestly to the public.	-0.17, $p = 0.341$
Steven is planning to vote for the government in the upcoming legislative election, even though he doesn’t support the government. He lives in fear of what will happen if he doesn’t vote for the government which normally wins more than 95% of votes in elections.	-0.43, $p = 0.029$

Table 3: **National Differences in Beliefs about Free and Fair Elections.** For each scenario, raters were asked to give a score ranging from 0 (not free at all) to 4 (completely free). The differences represent the average response for non-nationals minus nationals, and the p -values are calculated based on t-tests with equal variances.

be between “moderately free” and “very free” (with an average score of 2.5).

Although we cannot determine why Ugandan and non-Ugandan experts differed in their evaluations of this scenario and the other abstract scenarios, it is consistent with the idea that lived experience shapes how people – even experts – think about democracy in the abstract. For example, both Ugandan and non-Ugandan experts may have observed the “James” scenario first-hand, but may come to different conclusions about the extent to which

it indicates that free and fair elections exist or matters for the overall quality of democracy. National and non-national respondents' views of the vignettes supports the argument, made by Schaffer (2000) and others, that democracy means different things in different cultures. By asking experts to evaluate countries according to the V-Dem and (in our original survey) Freedom House criteria we may therefore be underestimating the differences in nationals' and non-nationals' *personal* assessments of how democratic countries are.

Our findings are potentially consistent with all three mechanisms we theorized: differences in information access, differences in ideas about democracy, or in-group bias. However, since the in-group bias mechanism predicts that national raters will consistently evaluate their countries more favorably than non-national raters, the more selective favorability we find is more consistent with the expectations of the other mechanisms.

Implications for Ratings' Authority

Thus far, we have provided evidence that expert raters of different nationalities assess democracy differently. Will raters' identities also matter for how GPIs are used?

GPIs are more likely to have authority when they come from trusted sources (Kelley and Simmons, 2019, 496). But where does that trust come from? Much of the literature can be read as implicitly suggesting that raters' nationality is beside the point. The rating entity – rather than the raters themselves – is usually emphasized. Especially relevant for our study is the idea that if an assessment comes from a recognized NGO, people may perceive it as high quality and legitimate (Honig and Weaver, 2019, 580). Supporting this logic, Nielson, Hyde and Kelley (2019) find that civil society actors are more likely to view election observers as legitimate if they are associated with prominent organizations and NGOs.

We argue, by contrast, that raters' nationality is also potentially significant for how ratings are used by policy practitioners. Ratings produced by fellow nationals might be more

trusted because they are perceived as more likely to reflect shared approaches and values. This hypothesized dynamic offers an explanation for why ratings entities are produced so overwhelmingly in North America and Western Europe (Kelley and Simmons, 2019, 497), which is where the greatest demand for GPIs has been located.

To examine how national identity affects ratings' authority, we recruited another sample of Ugandan and non-Ugandan elite users of democracy assessments in 2019. This study was also pre-registered (see SI §7). The study participants included current and former staffers of aid agencies, international NGOs working on issues of democracy and development, and others; this sampling approach reflects our understanding of the main non-academic users of democracy ratings as described in other studies (e.g., Büthe, 2012; Bush, 2017).²¹ We asked respondents to assess the trustworthiness of three democracy ratings: the one produced by national experts in our Uganda expert survey; the one produced by non-national experts in the same survey; and the one produced for Uganda by Freedom House. Focusing on these three types of ratings enabled us to explore how rater identity affects authority when a common methodology is used. After presenting respondents with descriptions of the three ratings in which only identity of the raters varied, we asked them to evaluate how trustworthy they were on a 10-point scale.²²

As expected, rater nationality affected perceptions of the ratings' trustworthiness. Non-national raters' assessments were significantly more trusted among non-nationals. As shown in Figure 6a, both the Freedom House score and the non-national experts' score were trusted significantly more than the Ugandan experts' score ($p < 0.001$ for both comparisons). By contrast, national raters' indicators had significantly more authority among national audiences. As Figure 6b shows, the Ugandan experts' score was trusted significantly more than

²¹This study was also pre-registered with Evidence in Governance and Politics as shown in SI §7.

²²In SI §6, we show results when we use responses to a question about which type of rating respondents would prefer to use in a professional setting.

the official Freedom House score ($p = 0.028$). It was also trusted more than the non-national experts' score, although this difference was not statistically significant at conventional levels ($p = 0.157$). Contrary to the expectation about the importance of rater branding, the official Freedom House score was not significantly more trusted than the non-branded international experts from our earlier survey among either national ($p = 0.441$) or non-national ($p = 0.226$) audiences.

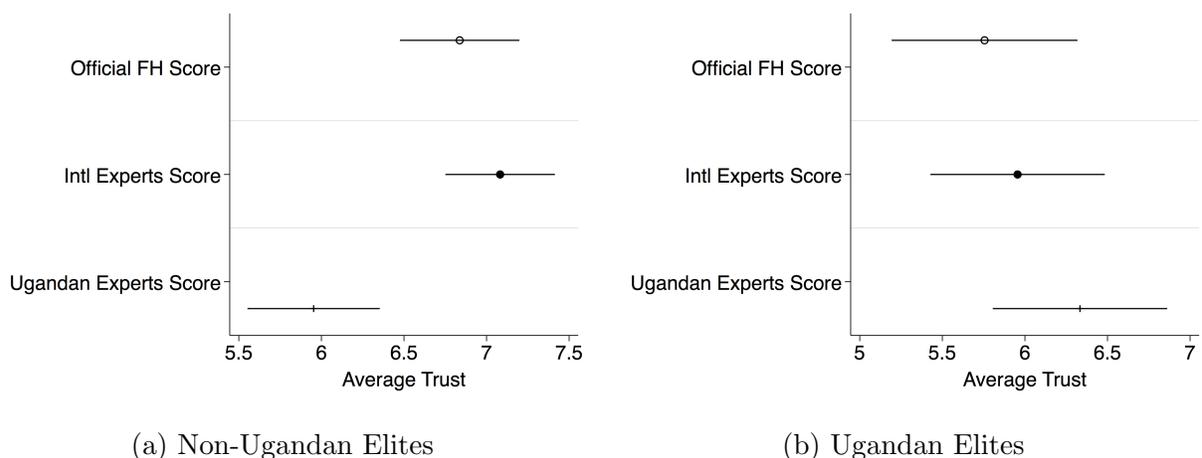


Figure 6: **Trust in Ratings by Rater Nationality.** This figure shows means and 95 percent confidence intervals. Trust is measured on a 10-point scale.

An open-ended follow-up question asked respondents about the reasons for their preferred rating. We provide illustrative responses associated with each type of preferred rating in Table 4. The most-commonly raised concern was about the ratings' accuracy, which was referenced by 29 percent of the respondents (for more details on coding see SI §6). Non-Uganda respondents often mentioned the potential bias or pressure Ugandan raters might face. This group also had a relatively favorable view of the official Freedom House score, and respondents who rated it highly also mentioned its reputation and global recognition as a rationale for their choice (9 percent of respondents). By contrast, those preferring the nationals' assessment tended to mention other accuracy considerations, such as better understanding of context, better information, and more experience of life “on the ground.”

These responses are quite consistent with our expectations about the perceived advantages of each type of rating.

Rationales for preferring non-national experts’ score:

–Assessing democracy needs to take into account several dimensions hence the need to consult more than one expert’s view. Also, being a national of the country being assessed may bring some analytical bias.

–I believe it is possible that Ugandan experts might be a bit “home-blind” and willing to have a more positive view of their country.

–I would be concerned about bias from the Ugandan experts.

–. . . there is less risk of political or other vested interests than the score based purely on Ugandan experts

Rationales for preferring the national experts’ score:

–The Ugandan experts are likely to understand the dynamics invoked and as well have an historical perspective

–I expect that the Ugandan experts score to be in better touch an understanding of the local context since they live it.

–Because their analysis is informed by experience as well information acquired from different sources. The others rely mostly on secondary information that is either embellished or not fully shared depending on the interests of their sources.

Rationales for preferring the official Freedom House score:

–I believe Freedom House has the best metric to measure and has the least “personal interest” in the score

–It is a source I have known and used for years; I trust its methodology and its “brand name” makes it easy to share information and have it accepted by other actors.

–Freedom House has more legitimacy.

Table 4: **Illustrative Responses about Preferred Democracy Score.** Survey respondents were invited to share their reasoning about which democracy assessment to use in a professional setting in an open-ended question.

Overall, these patterns suggest raters’ national identities contribute to ratings’ authority in significant ways, and that these effects vary with the audience. We highlight two further take-aways. First, non-Ugandan respondents could have spoken favorably about Ugandan ratings out of deference to shifting norms related to positionality and knowledge production. Aid officials, for example, increasingly speak about the importance of “local ownership” (Honig and Gulrajani, 2018), recognizing the pitfalls of international interventions that pay

insufficient attention to local voices (Autesserre, 2014). These dynamics have encouraged international actors to invest more in assessments with local (or peer-country) input (Swedlund, 2013, 453). Consequently, they might have been expected to show deference to the Ugandan experts' score. They did not.

Second, these findings suggest the possibility that externally-produced ratings are less trusted in the countries being rated. As such, the results may demonstrate some of the limits of criticism of country performance by international raters as a form of social shaming. Shaming is one important mechanism by which GPIs are thought to affect politics in the countries being evaluated (Kelley and Simmons, 2015). But if the ratings are not trusted by audiences there, then shaming may not be effective unless it is tied to material consequences.

Conclusion

Global performance indicators are influential and widely-used tools in real-world decision-making and in academic research. In this paper, we have shown that the national identity of expert raters called on to produce GPIs can matter for assessments of democracy as well as for their perceived authority.

We find that in the case of Uganda, national experts rate the country as substantially freer than non-national experts. This paper cannot establish what the “true” score for democracy is in Uganda, and thus it is not possible to say whether national experts are positively biased, non-national experts negatively biased, or both. Looking beyond the Uganda case, however, we also detect national differences among V-Dem raters.²³ While this pattern does not hold for all indicators or regions, it is present on numerous key outcomes, and especially for ratings of wealthy countries.

²³We hasten to emphasize that this analysis is not meant as a critique of V-Dem, which aggregates individual raters' scores through a sophisticated Bayesian model that can account for rater differences. In fact, it underscores the value of V-Dem's unique approach.

Together these results suggest the need to look more carefully at the role rater identity plays in democracy assessments and GPIs more generally. The institutions that produce GPIs regularly reassess their methodologies, and our interviews suggest that rater nationality is among their concerns. With that in mind, we highlight three implications for producers and users of democracy ratings. First, our findings demonstrate that is important to recognize the way that rater identity shapes the content of such assessments. Doing so would follow Colgan (2019*b*)’s advice to be attentive to the potential for national bias, similar to other forms of measurement bias that are already familiar to political scientists. V-Dem’s collection of meta-data on experts offers a best practice for other datasets and GPIs that is rarely adopted. At the very least, transparency about the question of “who rates” may be important for the legitimacy and reliability of these indicators going forward.

Second, our findings raise questions about the validity and reliability of existing GPIs, and the real-world consequences in terms of policy and knowledge production resulting from potentially systematic bias in assessments. If rater identity *does* matter for assessments, at least for some indicators in some places, might we – by virtue of systematic differences in the representation of nationals in GPI production across countries and regions – come to incorrect conclusions in either the levels of indicators in particular countries, or in the ranking of countries according to these indicators? How might different ratings have affected decisions about foreign aid, or the findings of research relying on these indicators?

Third, our findings raise normative questions about who *ought* to evaluate democracy in a given country. Our findings from the V-Dem data suggest that, first, Europe is one of the regions with the highest percentage of nationals participating in ratings and on second, that the gap between national and non-national raters seems to increase in wealthier countries. Systematic overrepresentation of wealthy countries among the raters and among the organizations producing GPIs could thus introduce significant bias in favor of wealthier, Global North countries. We think that such differences are even more likely for other GPIs,

few of which make explicit commitments to include national raters in the way V-Dem has. An implication is that how we perceive democracy in some regions may be more affected by national bias than others.

A few recent efforts have tried to include more “local” voices in the content and creation of governance indicators. These initiatives suggest that the problem of representation is not necessarily easily rectified. For example, Swedlund (2013, 465) notes that a Joint Governance Assessment in Rwanda faced considerable difficulties including “alternative expectations and low levels of domestic institutional capacity,” with external and domestic assessors coming to different and conflicting conclusions on key measures. The possibility of domestic political interests entering the assessment framework is a real concern. Nevertheless, the solution should not be to simply maintain the status quo but rather to continue working to find ways to include a wider variety of experts in the assessment process.

Finally, turning to the topic of GPIs’ authority, would national assessments for the U.S. and Germany be as distrusted as Ugandan assessments of Uganda are by international audiences? The existence of respected, high-profile assessments for democracy in the U.S. such as Bright Line Watch and the Authoritarian Warning Survey, which are dominated by American experts’ evaluations of American democracy, seems to provide prima facie evidence that they are not. Are concerns about national rater bias heightened in less democratic contexts, or are other factors at play? It could be that rater identity matters less in cases where there is greater consensus about the regime type or the regime type in which raters reside is held constant. For example, it could be that national and non-national experts from consolidated democracies assess the extent of democracy more similarly, and thus rater nationality will matter less. We see these as questions and directions that could further our understanding of how GPI are created and the extent to which biases may feature in these influential tools in global governance.

References

- Andersson, Staffan and Paul M. Heywood. 2009. "The Politics of Perception: Use and Abuse of Transparency International's Approach to Measuring Corruption." *Political Studies* 57(4):746–767.
- Autesserre, Sèverine. 2014. *Peaceland: Conflict Resolution and the Everyday Politics of International Intervention*. Problems of International Politics Series. New York: Cambridge University Press.
- Barnett, Michael N. and Martha Finnemore. 2004. *Rules for the World: International Organizations in Global Politics*. Ithaca, N.Y.: Cornell University Press.
- Besley, Timothy. 2015. "Law, Regulation, and the Business Climate: The Nature and Influence of the World Bank Doing Business Project." *Journal of Economic Perspectives* 29:99–120.
- Bilodeau, Antoine. 2014. "Is democracy the only game in town? Tension between immigrants' democratic desires and authoritarian imprints." *Democratization* 21(2):359–381.
- Bisbee, James H., James R. Hollyer, B. Peter Rosendorff and James Raymond Vreeland. 2019. "The Millennium Development Goals and Education: Accountability and Substitution in Global Assessment." *International Organization* 73(3):547–578.
- Bollen, Kenneth A. and Pamela Paxton. 2000. "Subjective Measures of Liberal Democracy." *Comparative Political Studies* 33(2):58–86.
- Bradley, Christopher G. 2015. International Organizations and the Production of Indicators: The Case of Freedom House. In *The Quiet Power of Indicators: Measuring Governance, Corruption, and the Rule of Law*, ed. Sally Engle Merry, Kevin E. Davis and Benedict Kingsbury. Cambridge: Cambridge University Press pp. 27–74.
- Bratton, Michael. 2010. "Anchoring the "D-Word" in Comparative Survey Research." *Journal of Democracy* 21(4):106–113.
- Bratton, Michael, Robert Mattes and Emmanuel Gyimah-Boadi. 2005. *Public Opinion, Democracy, and Market Reform in Africa*. Cambridge: Cambridge University Press.
- Brooks, Sarah M., Raphael Cunha and Layna Mosley. 2015. "Categories, Creditworthiness, and Contagion: How Investors' Shortcuts Affect Sovereign Debt Markets." *International Studies Quarterly* 59(3):587–601.
- Bush, Sarah Sunn. 2017. "The Politics of Rating Freedom: Ideological Affinity, Private Authority, and the Freedom in the World Ratings." *Perspectives on Politics* 15(3):711–731.

- Büthe, Tim. 2012. Beyond Supply and Demand: A *Political-Economic* Conceptual Model. In *Governance by Indicators: Global Power through Classification and Rankings*, ed. Kevin Davis, Angelina Fisher, Benedict Kingsbury and Sally Engle Merry. New York: Oxford University Press pp. 29–51.
- Carothers, Thomas. 1997. “The Observers Observed.” *Journal of Democracy* 8(3):17–31.
- Cheng, Christine and Alison Brett. 2019. “How Cognitive Frameworks Shape the American Approach to International Relations and Security Studies.” *Journal of Global Security Studies* 4(3):321–344.
- Colgan, Jeff D. 2019a. “American Bias in Global Security Studies Data.” *Journal of Global Security Studies* 4(3):358–371.
- Colgan, Jeff D. 2019b. “American Perspectives and Blind Spots on World Politics.” *Journal of Global Security Studies* 4(3):300–309.
- Cooley, Alexander. 2015. The Emerging Politics of International Rankings and Ratings: A Framework for Analysis. In *Ranking the World: The Politics of International Rankings and Ratings*, ed. Alexander Cooley and Jack Snyder. New York: Cambridge University Press pp. 1–38.
- Coppedge, Michael, John Gerring, Adam N. Glynn, Carl Henrik Knutsen, Staffan I. Lindberg, Daniel Pemstein, Brigitte Seim, Svend-Erik Skaaning and Jan Teorell. 2019. *Varieties of Democracy: Measuring Two Centuries of Political Change*. Cambridge: Cambridge University Press.
- Coppedge, Michael, John Gerring, David Altman, Michael Bernhard, Steven Fish, Allen Hicken, Matthew Kroenig, Staffan I. Lindberg, Kelly McMann, Pamela Paxton, Holli A. Semetko, Svend-Erik Skaaning, Jeffrey Staton and Jan Teorell. 2011. “Conceptualizing and Measuring Democracy: A New Approach.” *Perspectives on Politics* 9(2):247–267.
- Cross, Mai’a K. Davis. 2013. “Rethinking Epistemic Communities Twenty Years Later.” *Review of International Studies* 39(1):137–160.
- Dietrich, Simone. 2016. “Donor Political Economies and the Pursuit of Aid Effectiveness.” *International Organization* 70(1):65–102.
- Dolan, Lindsay R. 2017. “Labeling Laggards and Leaders: International Organizations and the Politics of Defining Development.” Working Paper.
- Doshi, Rush, Judith G. Kelley and Beth A. Simmons. 2019. “The Power of Ranking: The Ease of Doing Business Indicator and Global Regulatory Behavior.” *International Organization* 73(3):611–643.

- Engel, Eliot T. 2020. "Letter to Secretary of State Mike Pompeo and Treasury Secretary Steven Mnuchin." Foreign Affairs Committee of the U.S. House of Representatives. Available at https://foreignaffairs.house.gov/_cache/files/7/d/7d49f280-b8a5-4789-9d3a-dd8bd6310b3d/A726593A9B7DEEC19C35AD973FE84BE6.12-9-2020-ele-letter-to-pompeo-mnuchin-uganda.pdf.
- English, Cynthia and Lee Becker. 2013. "Majorities in Most Countries Perceive Their Media as Free." Gallup.com, May 3. Available at <https://news.gallup.com/poll/162179/majorities-countries-perceive-media-free.aspx>.
- Espeland, Wendy Nelson and Michael. 2007. "Rankings and Reactivity: How Public Measures Recreate Social Worlds." *American Journal of Sociology* 113(1):1–40.
- Findley, Michael G., Adam S. Harris, Helen V. Milner and Daniel Nielson. 2017. "Who Controls Foreign Aid? Elite versus Public Perceptions of Donor Influence in Aid-Dependent Uganda." *International Organization* 71(4):633–663.
- Fuchs, Andreas and Kai Gehring. 2017. "The Home Bias in Sovereign Ratings." *Journal of the European Economic Association* 15(6):1386–1423.
- Giannone, Diego. 2010. "Political and Ideological Aspects in the Measurement of Democracy: The Freedom House Case." *Democratization* 17(1):68–97.
- Girod, Desha M., Stephen D. Krasner and Kathryn Stoner-Weiss. 2009. Governance and Foreign Assistance: the Imperfect Translation of Ideas into Outcomes. In *Promoting Democracy and the Rule of Law: American and European Strategies*, ed. Amichai Magen, Thomas Risse and Michael A. McFaul. London: Palgrave Macmillan pp. 61–92.
- Gray, Julia and Jonathan Slapin. 2012. "How Effective Are Preferential Trade Agreements? Ask the Experts." *Review of International Organizations* 7(2):309333.
- Gray, Julia and Raymond P. Hicks. 2014. "Reputations, Perceptions, and International Economic Agreements." *International Interactions* 40(3):325–349.
- Green, Jessica F. 2013. *Rethinking Private Authority: Agents and Entrepreneurs in Global Environmental Governance*. Princeton: Princeton University Press.
- Gunitsky, Seva. 2015. Competing Measures of Democracy in the Former Soviet Republics. In *Ranking the World: The Politics of International Rankings and Ratings*, ed. Alexander Cooley and Jack Snyder. New York: Cambridge University Press pp. 112–150.
- Haas, Peter M. 1992. "Epistemic Communities and International Policy Coordination." *International Organization* 46(1):1–35.
- Hobson, Christopher and Milja Kurki. 2012. Introduction. In *The Conceptual Politics of Democracy Promotion*, ed. Christopher Hobson and Milja Kurki. London: Routledge pp. 1–15.

- Honig, Dan and Catherine Weaver. 2019. "A Race to the Top? The Aid Transparency Index and the Social Power of Global Performance Indicators." *International Organization* 73(3):579–610.
- Honig, Daniel and Nilima Gulrajani. 2018. "Making Good on Donors' Desire to Do Development Differently." *Third World Quarterly* 39(1):68–84.
- Hyde, Susan D. 2011. *The Pseudo-Democrat's Dilemma: Why Election Observation Became an International Norm*. Ithaca, N.Y.: Cornell University Press.
- Hyde, Susan D. 2012. Why Believe International Election Monitors? In *The Credibility of Transnational NGOs: When Virtue is Not Enough*, ed. Peter A. Gourevitch, David A. Lake and Janice Gross Stein. Cambridge: Cambridge University Press pp. 37–61.
- Just, Aida. 2019. "Political Regimes and Immigrant Party Preferences." *Comparative Political Studies* 52(5):651–686.
- Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi. 2010. "The Worldwide Governance Indicators: Methodology and Analytical Issues." World Bank Policy Research Working Paper No. 5430. World Bank, Washington, D.C.
- Kelley, Judith G. and Beth A. Simmons. 2015. "Politics by Number: Indicators as Social Pressure in International Relations." *American Journal of Political Science* 59(1):55–70.
- Kelley, Judith G. and Beth A. Simmons. 2019. "Introduction: The Power of Global Performance Indicators." *International Organization* 73(3):491–510.
- Kerner, Andrew, Morten Jerven and Alison Beatty. 2017. "Does It Pay to be Poor? Testing for Systematically Underreported GNI Estimates." *Review of International Organizations* 12(1):1–38.
- King, Gary, Christopher J.L. Murray, Joshua A. Salomon and Ajay Tandon. 2004. "Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research." *American Political Science Review* 90(1):191–207.
- Logan, Carolyn and Robert Mattes. 2012. "Democratising the Measurement of Democratic Quality: Public Attitude Data and the Evaluation of African Political Regimes." *European Political Science* 11(4):469–491.
- Lührmann, Anna, Marcus Tannenbergh and Staffan I. Lindberg. 2018. "Regimes of the World (RoW): Opening New Avenues for the Comparative Study of Political Regimes." *Politics and Governance* 6(1):1–18.
- Morse, Julia C. 2019. "Blacklists, Market Enforcement, and the Global Regime to Combat Terrorist Financing." *International Organization* 73(3):511–545.

- Munck, Geraldo L. and Jay Verkuilen. 2002. "Conceptualizing and Measuring Democracy." *Comparative Political Studies* 35(1):5–34.
- Mutz, Diana C. and Eunji Kim. 2017. "The Impact of In-group Favoritism on Trade Preferences." *International Organization* 71(4):827–850.
- Nielson, Daniel L., Susan D. Hyde and Judith G. Kelley. 2019. "The Elusive Sources of Legitimacy Beliefs: Civil Society Views of International Election Observers." *The Review of International Organizations* 14(4):685–715.
- Norris, Pippa. 2013. "Does the World Agree about Standards of Electoral Integrity? Evidence for the Diffusion of Global Norms." *Electoral Studies* 32(4):576–588.
- Oren, Ido. 1995. "The Subjectivity of the "Democratic" Peace: Changing U.S. Perceptions of Imperial Germany." *International Security* 20:147–184.
- Pietsch, Juliet. 2015. "Authoritarian Durability: Public Opinion towards Democracy in Southeast Asia." *Journal of Elections, Public Opinion and Parties* 25(1):31–46.
- Razafindrakoto, Mireille and François Roubaud. 2010. "Are International Databases on Corruption Reliable? A Comparison of Expert Opinion Surveys and Household Surveys in Sub-Saharan Africa." *World Development* 38(8):1057–1069.
- Schaffer, Frederic Charles. 2000. *Democracy in Translation: Understanding Politics in an Unfamiliar Culture*. Ithaca, NY: Cornell University Press.
- Seabrooke, Leonard and Duncan Wigan. 2015. "Powering Ideas Through Expertise: Professionals in Global Tax Battles." *Journal of European Public Policy* 23(3):357–374.
- Sears, David O. and Christia Brown. 2003. Childhood and Adult Political Development. In *The Oxford Handbook of Political Psychology*.
- Steiner, Nils D. 2016. "Comparing Freedom House Democracy Scores to Alternative Indices and Testing for Political Bias: Are U.S. Allies Rated as More Democratic by Freedom House?" *Journal of Comparative Policy Analysis* 18(4):329–349.
- Swedlund, Haley J. 2013. "The Domestication of Governance Assessments: Evidence from the Rwandan Joint Governance Assessment." *Conflict, Security & Development* 13(4):449–470.
- Swedlund, Haley J. 2017. "Can Foreign Aid Donors Credibly Threaten to Suspend Aid? Evidence from a Cross-national Survey of Donor Officials." *Review of International Political Economy* 24(3):454–496.
- Terman, Rochelle. 2017. "Islamophobia and Media Portrayals of Muslim Women: A Computational Text Analysis of US News Coverage." *International Studies Quarterly* 61(3):489–502.