

National Identity and Democracy Ratings

Sarah Sunn Bush¹  and
Melina R. Platas² 

Comparative Political Studies
2025, Vol. 58(6) 1141–1170
© The Author(s) 2024



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00104140241259457
journals.sagepub.com/home/cps



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, this article suggests raters' identities may shape their assessments. It specifically examines how national identity shapes democracy ratings. Two data sources—an original survey of experts on Uganda and the Varieties of Democracy Institute—reveal significant differences in the ratings provided by national and non-national experts. In most cases, ratings by nationals are more positive. This article explores three potential reasons for the difference, finding some support for each: national differences in information access and consumption, national differences in conceptions of democracy, and in-group-out-group bias. The findings have implications for our understanding of global performance indicators, which are overwhelmingly a product of Global North organizations.

Keywords

African politics, democratization and regime change, human rights, non-democratic regimes, quality of democracy, global performance indicators

¹University of Pennsylvania, Philadelphia, PA, USA

²NYU Abu Dhabi, United Arab Emirates

Corresponding Author:

Sarah Sunn Bush, Department of Political Science, University of Pennsylvania, 133 S. 36th Street, Philadelphia, PA 19104-6243, USA.

Email: ssbush@upenn.edu

Data Availability Statement included at the end of the article

The recent growth of global performance indicators (GPIs) is remarkable. In the late 1990s, only twenty GPIs existed; by the mid-2010s there were eight times as many (Cooley, 2015; Kelley & Simmons, 2019, p. 493). Almost all (95%) GPIs are produced by organizations headquartered in the Global North, particularly Europe and North America (Kelley & Simmons, 2019, p. 497). Nearly half are produced in the United States alone. Thus, all countries are assessed, but actors in very few countries are assessors. Although some individuals from the Global South are raters, they remain under-represented in the production of GPIs.

We argue that this under-representation matters because rater identity shapes expert assessments. Building on prior work (Colgan, 2019a), we focus on the relationship between rater nationality and democracy ratings. Rater nationality may affect experts' democracy ratings through three mechanisms: differences in information access and consumption; differences in conceptions of democracy; and in-group–out-group bias. These mechanisms are not mutually exclusive. All suggest reasons why national raters will reach different—and, with the in-group–out-group bias mechanism, more favorable—conclusions than non-national raters about how democratic countries are.

Although democracy ratings are only one type of GPI, they are of substantive importance. They are widely used by scholars as well as by non-academics in foreign aid policy, non-governmental organization (NGO) advocacy, and risk assessment (Girod et al., 2009; Kaufmann et al., 2010, p. 20; Bradley, 2015, p. 53). They therefore have the potential to alter real-world decisions, particularly regarding investment, aid, and political engagement. Democracy ratings are also typical of GPIs in their reliance on expert judgment.¹ They are usually made by country or regional experts who—pertinent for our inquiry—vary in whether they are nationals of the country being assessed.

We use two data sources to examine the relationship between rater nationality and democracy ratings. The first is an original expert survey. The survey assessed democracy in Uganda, a case we chose because—as a competitive authoritarian regime—it is a most-likely case for expert disagreement about its regime type. Our survey recruited Ugandan and non-Ugandan governance experts to assess Uganda using a methodology modeled on what is used by some official democracy ratings. We find that Ugandan experts rated Uganda as significantly more democratic than non-Ugandan experts.

The second data source is the Varieties of Democracy (V-Dem) Institute (Coppedge et al., 2011), which surveys both national and non-national experts to create a leading measure of democracy. Analyzing the V-Dem data enables us to consider whether the findings from our original survey generalize. Examining the assessments of nearly 2300 raters across 180 countries and

more than 70 V-Dem measures, we find that national raters offered significantly different, and mostly more favorable, assessments than non-national raters.

Using both data sources, we also find evidence in support of each of the three proposed mechanisms linking nationality to assessments. For example, we document systematic differences in experts' information sources and views about democracy in the abstract when comparing nationals to non-nationals. Although the direction of the nationality difference is mostly consistent with the in-group–out-group mechanism, nationals are sometimes more critical than non-nationals. Thus, the key finding is that national differences in assessments are possible and even likely. The size of the difference is large enough to have real-world implications for countries' populations, as we show for Uganda, which relies on aid that is conditional on democracy GPIs.

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

Finally, this study contributes to our understanding of the reliability and validity of GPIs. Studies have documented methodological problems in many prominent indicators, including the *Freedom in the World* reports produced by Freedom House (Gunitsky, 2015), the Corruptions Perceptions Index produced by Transparency International (Andersson & Heywood, 2009), the Trafficking in Persons report produced by the U.S. State Department (Harmon et al., 2022), and credit ratings such as Fitch, Moody's, and Standard and Poor's (Fuchs & Gehring, 2017). The national identity of raters is an additional consideration. Beyond the recognized bias in favor of the Global North in the production of GPIs, there is evidence of national bias in raters' coding decisions, and we add to it (Colgan, 2019a). We show below that the gap

between national and non-national raters in democracy assessments is largest among countries in the Global North, suggesting that the extent of national bias in prominent GPIs may vary by country. Our findings therefore have important implications for GPI producers, and we take this theme up in our conclusion.

Why Rater Identity Matters

Ratings and rankings are produced by *people*. Yet the literature on GPIs generally does not theorize or examine the effect of variation in those individuals' characteristics.² More attention has been paid to the institutions where raters work. For example, inter-governmental organizations (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 & Finnemore, 2004, pp. 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, p. 716), or the World Bank economists who worked on the *Doing Business* report (Besley, 2015, pp. 101–102). In other cases, experts are non-academic professionals working in IGOs, NGOs, government agencies, and the media. In general, experts 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, p. 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, experts may share common ideas about the concepts being rated. After all, shared causal and principled beliefs are tied to the development of professional knowledge (Seabrooke & Wigan, 2015). Second, experts may have access to and rely on similar information about country performance. Supporting this logic, Razafindrakoto and Roubaud (2010, pp. 1063–1066) found that experts show some signs of group think—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. Coder trainings can also help. 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: *The country assessments of national and non-national expert raters will differ.*

Our starting point is 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, raters often disagree on how to assess countries that are neither consolidated democracies nor autocracies. 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 post-Soviet hybrid regimes of that era. Such differences may be the result of divergent conceptions of democracy and measurement strategies. But different raters could interpret 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, 2019b](#), p. 300) and, more broadly, to use distinct cognitive frames ([Cheng & Brettle, 2019](#)). By national identity, we follow other research in referring primarily to the country where an expert was born and lives ([Colgan, 2019b](#), p. 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 and consumption; differences in conceptions of democracy; and in-group–out-group bias. Although distinct in theory, these mechanisms may be complementary in practice, and all predict that national and non-national raters will reach different ratings. The in-group–out-group bias mechanism further predicts that national raters will rate their own country more favorably. 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 national media than non-nationals.⁴ Although some sources are widely available, content from many smaller newspaper, television, and radio outlets is often not available online and may be inaccessible to non-national experts for other reasons (e.g., language). These sources sometimes cover different events than international media (e.g., [Clarke, 2023](#)), and they may frame the same events differently.

For example, the American media tends to cover countries in ways that reflect U.S. foreign policy priorities and stereotypes (Terman, 2017). These differences matter because the information experts have access to shapes their assessments of country performance on issues such as human rights (Clark & Sikkink, 2013; Fariss, 2014).

In addition to differences in information access, we expect there to be national differences in information consumption and exposure. Nationality shapes individuals' social networks and thus the content they are passively exposed to through social media and interpersonal discussion. It may also shape which sources an expert finds valuable. Given these differences, non-national raters may be more likely to rely on other GPIs when making assessments.

The information mechanism implies that national differences in information consumption, knowledge, and country assessments will exist. In terms of the latter, the information mechanism does not predict the direction of difference. It is also not obvious whether nationals or non-nationals will be more likely to reach the "correct" assessment of a country's performance. One view is that nationals are in a better position to assess countries given their better access to local information; this perspective is taken by V-Dem, which includes mostly national raters (Coppedge et al., 2019, p. 11). Another is that non-nationals are better able to take a comparative perspective by having better access to and consumption of information about a range of countries.

Second, nationals and non-nationals may evaluate the same country differently because they see the world differently. Many GPIs concern contested ideals. Both academics and policymakers disagree about how to define "democracy" beyond the minimum of contested elections. Insofar as a rating system provides detailed rules about how to evaluate countries on various sub-indicators, these disagreements matter less, but the challenge of eliminating all subjectivity in expert ratings is significant (Knutsen et al., 2024).

It is plausible that individuals' lived experiences and values will prompt them to interpret the concepts being rated in ways that reflect their national identities. Political orientation develops early and can be sticky (Sears & Brown, 2013). For example, immigrants' ideas about governance vary systematically with the type of political regime in their country of origin (Bilodeau, 2014; Just, 2019). Thus, variation in experts' nationalities may influence their assessments of country performance. We draw out observable implications about the distinction between nationals and non-nationals in particular given our focus. Again, insofar as ratings ask experts to evaluate specific sub-indicators of democratic performance, these value differences may be less relevant. Yet it is difficult to eliminate all ambiguity from many ratings, and raters' views on one dimension of country performance may spill over into their views of another dimension (Brooks et al., 2015; Gray & Hicks, 2014).

The second mechanism implies that national and non-national experts will evaluate the same country differently. 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–out-group bias. Whereas the first two mechanisms imply that nationals and non-nationals will 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 will provide more negative assessments as an expression of negativity towards a national out-group. To be clear, we do not take a position about which group of raters (nationals or non-nationals) 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 & Kim, 2017](#)).⁵

Since democracy is generally perceived as a normatively-desirable trait, the logic of in-group favoritism and out-group derogation implies that in-group experts will view their own country as more democratic than out-group experts. This logic further suggests that experts will view out-group countries less favorably in general and as less democratic in particular. Whereas the other two mechanisms do not necessarily predict that nationals and non-nationals will reach different assessments across the board (e.g., on all sub-components of democracy), the in-group favoritism mechanism expects more consistent differences. Although such a bias is surprising if we believe experts form an epistemic community with internal cohesion and professionalism, this type of behavior is more plausible if we view expert raters to be less cohesive as a group, perhaps because they have been trained at different institutions, meet infrequently, and lack a common culture ([Cross, 2013](#), p. 150).

Research Design

To test our hypothesis against the null, we use two types of data.⁶ 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 evaluations of national and non-national experts across all countries over a range of years.

Each data source has advantages. Since we conducted the Uganda survey ourselves, we could include questions designed to examine the three mechanisms 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 more common in recent years (Lührmann et al., 2018, p. 8).

Our second data source is V-Dem. Here, we can examine the relationship between rater nationality and democracy assessments across many indicators. This analysis addresses two concerns about the external validity of the Uganda survey. The first is whether our findings generalize to other countries, time periods, and a larger sample of experts. The second is whether the findings would be similar in a real-world rating, although we attempted to approximate one with our survey. That our findings about national differences in the V-Dem data corroborate our findings from the Uganda survey provides further support that nationality matters for expert assessments.

Uganda Expert Survey

Our original survey builds on other studies that seek to approximate expert ratings exercises through elite surveys (Arnon et al., 2023; Razafindrakoto & Roubaud, 2010). We recruited a sample of experts on Ugandan politics to participate in an online survey in 2019. We pre-registered our analysis plans with the Evidence in Governance and Politics repository.⁷ First we provide our logic of case selection and background information and then describe the sample.

Case Selection and Background

Uganda is classified as a competitive authoritarian regime by most GPIs. For example, 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” as of 2024. 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 as of 2018, at the border between the Polity categories “closed anocracy” and “open anocracy.” We chose to study a competitive authoritarian regime because countries that fall into the middle of the democracy–autocracy spectrum are traditionally the ones about which experts are most likely to disagree, as Gunitsky (2015) has shown in the case of the former Soviet republics.⁸

Uganda has been governed by the National Resistance Movement (NRM) and President Yoweri Museveni for nearly 40 years. During that time, 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. Elections for parliament and local government are competitive, and many incumbents lose their seats in each election. At the same time, there is evidence that constraints on the executive have declined. For example, the parliament voted to eliminate presidential term limits in 2005 and age limits in 2018. These changes allowed Museveni to continue participating in elections. Presidential elections have some competition, with opposition candidates garnering up to 40% of the vote, but Museveni has continued to win. 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. These features of Uganda’s regime make it a likely case for expert disagreement.⁹

Uganda is also precisely the type of country for which democracy ratings matter. It has relied on foreign aid for as much as 70% of government expenditures since the 1990s (Findley et al., 2017, p. 642). Perceptions about democracy inform donors’ aid decisions and other benefits (Hyde, 2011, Ch. 3; Bush & Zetterberg, 2021). In 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.

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. 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 academia, NGOs engaged in democracy promotion, and think tanks.¹¹ We targeted individuals with similar backgrounds. In total, 220 experts were invited via email, of whom 118 answered at least some questions for a 54% response rate. This sample size and response rate compare favorably with related expert surveys.¹²

To identify non-Ugandan experts, we used several methods. First, we searched the websites of top academic departments in political science and

African studies, using “Uganda” as a keyword. Second, we identified experts from the United States, Canada, and Europe who authored or co-authored works on Uganda by consulting programs from American Political Science Association meetings and using Google Scholar. Individuals entered the sample if they published multiple works on Uganda in the past ten years or were policy professionals whose work portfolios involved Uganda. Our interviews suggest this process closely resembles GPIs’ recruitment processes, with the exception that we could not mimic social network dynamics (e.g., when last year’s expert suggests a colleague to replace her).

To identify Ugandan experts, we focused on individuals working in several sectors who are knowledgeable about Ugandan politics (for more details, see SI §3). First, we included lecturers and faculty at the political science department (or equivalent) of four top Ugandan universities. Second, we included the senior staff of major governance-related NGOs and other important policy analysts based in Uganda. Third, 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.

The people in our sample reported considerable knowledge about Uganda, as expected. Seventy percent reported following Ugandan politics every day, and a further 22% reported following it a few times a week. Eighty-seven percent said that they were “somewhat” or “very” certain in their responses to questions about democracy in Uganda. We exclude respondents from 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 (see SI §3). The non-Ugandan participants were more likely to be women, work in academia, and possess doctoral degrees. They were also farther left ideologically. Although it is not obvious to us whether or how these background characteristics would 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 survey was completed anonymously using the Qualtrics platform. We doubt that many respondents would have felt concerned about government reprisals or social desirability in what was (truthfully) described as an academic study 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 Freedom House.¹⁴ *Freedom in the World* provides 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). The Freedom House case can also inform our understanding of other expert surveys of democracy (including Polity, V-Dem, and the Economist Intelligence Unit) and other GPIs that rely on this method (such as the Corruption Perceptions Index and Ease of Doing Business index).

The Freedom House questionnaire (see SI §4) 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 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 computed overall scores for political rights and civil liberties based on respondents’ answers.

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 has asked country experts to take an online survey for US\$25 compensation.¹⁵

Second, raters’ decisions were not debated by others. Freedom House accomplishes this task through ratings review meetings at the regional level. We may overestimate national differences by skipping this review step. We return in the conclusion to this point in our discussion of how GPIs do and should address the potential for national bias.

Main Findings

To test whether ratings differ by rater national identity, we first classify respondents as Ugandan or non-Ugandan based on their reported citizenship. In SI §3, we code the national in-group on the basis of residency as a robustness check. 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 ($p = .009$). Figure 1 shows the distribution of scores across

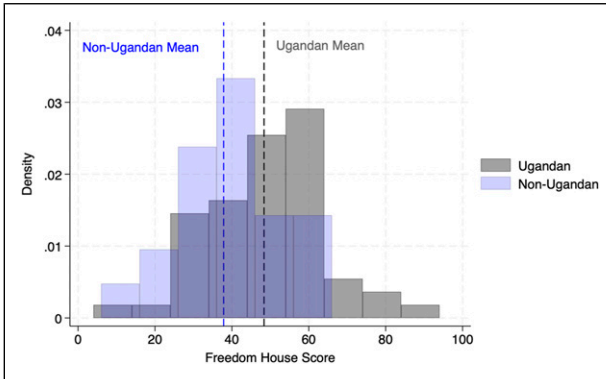


Figure 1. The Distribution of Democracy Scores by Rater Nationality. This figure shows experts' raw scores, with a score of 100 indicating greatest democracy. Difference = 10.51, $p = .006$ according to a two-tailed t test with equal variances.

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 1](#) 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% increase.¹⁶ In fact, the score provided by non-nationals was on the border of “not free” and “partly free,” whereas the nationals’ score was comfortably within the “partly free” category. The numerical scores are particularly important for some GPI users. For example, the Millennium Challenge Corporation, a U.S. foreign assistance program, uses twenty-five as the civil liberties’ cut off; non-Ugandan experts in our survey barely thought the country passed with an average score of twenty-six, compared to thirty-four for Ugandan experts.¹⁷

Freedom House groups questions into two categories: political rights and civil liberties. We examine national differences in both sets of questions, as well as the overall score, using ordinary least squares (OLS) regressions. For this analysis, we rescaled all outcome measures to be between zero and one. We also include covariates: rater demographic characteristics (gender, political ideology, and profession) and knowledge about Uganda (frequency of following Ugandan politics, certainty in responses, and number of Ugandan media sources followed).¹⁸ Nationality is a complex “treatment” that likely involves additional differences beyond what we can or did measure; this complexity is part of why we find the 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 2, Ugandan raters provided higher scores for both political rights and civil liberties, though the relationship is more consistently statistically significant for civil liberties. Later, we discuss some of the results on individual sub-indicators.

As a robustness check, we re-run the analysis after removing respondents with scores which could be considered outliers, that is, in the bottom 5% and top 95% of the distribution of raw scores (see SI §3). Our results are robust and in fact slightly more statistically significant.

V-Dem Democracy Assessments

To what extent do our findings from Uganda 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 V-Dem. For most democracy GPIs, including Freedom House, information about coders’ nationalities is either not maintained or inaccessible. By contrast, V-Dem gathers such information about its raters, alongside other forms of meta-data such as timestamps (Weidmann, 2024). 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 (2019b), p. 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.

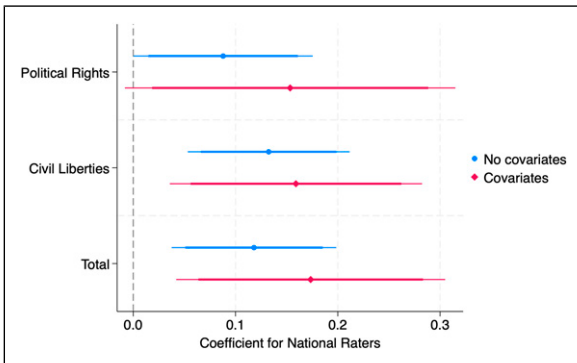


Figure 2. The Relationship between Rater Nationality and Democracy Assessments. The figure displays coefficient estimates with 95 and 90% confidence intervals based on OLS regressions of democracy assessments on an indicator for the respondent’s national identity. The covariates included in this analysis are rater gender, left–right political ideology, frequency of following Ugandan politics, profession, certainty in responses, and number of Ugandan media sources followed. See SI §3 for a table.

Description of the Nationality Data

V-Dem invites expert raters 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 questionnaire which queries their nationality and residency. The V-Dem data to which we have access covers nearly 2300 raters across 180 countries. Although most V-Dem experts are nationals in every world region by design, European countries are most likely to be assessed by nationals, and countries in the Middle East and North Africa and Oceania regions are least likely to be assessed by them (see SI §5). Out of consideration for raters' privacy, V-Dem does not share information on where the rater was born or resides if not in the rated country.¹⁹ Although V-Dem also gathers other information about raters' identities (e.g., age, gender, and education), we similarly could not access this information due to privacy concerns. As such, we cannot assess the extent to which national raters for a given country-year differ from non-national raters in terms of other characteristics.

Main Findings

We analyze two types of outcome variables from V-Dem. First, as part of the post-survey questionnaire, all raters are asked to provide an assessment of democracy in the main country they rated as of 2012. The instructions 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 call this 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 to aggregate across respondents and indicators. Yet the responses to this 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](#). In the baseline models, national raters assessed their main country of expertise to be significantly more democratic than non-national raters; the difference is about seven points on the 100-point scale in Model 1, or a more than 11% increase. The difference is even larger when considering residents versus non-residents: a 13-point or

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.

	(1)	(2)	(3)	(4)	(5)	(6)
	National	Resident	National	Resident	National	Resident
Coder is a national	6.857*** (1.424)		1.938 (1.308)		-0.489 (1.109)	
Coder is a resident		13.383*** (1.309)		7.576*** (1.196)		0.382 (0.945)
Constant	60.509*** (1.241)	57.011*** (1.119)	52.625*** (1.610)	49.701*** (1.538)	68.104*** (4.412)	67.481*** (4.449)
Region FE	No	No	Yes	Yes	No	No
Country FE	No	No	No	No	Yes	Yes
N	1774	1771	1774	1771	1774	1771
R ²	0.014	0.062	0.256	0.269	0.669	0.669

23% increase. When we introduce regional fixed effects (FEs) in the second set of models to account for regional variation in reliance on national raters, there is still a significant national difference when comparing residents and non-residents, although the size of the difference is almost cut in half. The inclusion of country fixed effects in the final set of models reduces the coefficients and statistical significance further. 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.

Second, as part of the main V-Dem questionnaire, raters were asked to assess many indicators. We focus on the indicators that contribute to at least one of five V-Dem democracy indices (electoral, liberal, participatory, deliberative, or egalitarian) and therefore bear on assessments of democracy. 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 our analysis of the 71 V-Dem indicators; SI §5 contains the full results. There were significant national differences on 80% of indicators. National raters were usually more favorable, similar to the direction of difference in both the democracy scores and the Uganda expert survey. But in 20% of cases, they were more critical. The overall picture is thus one in which national differences among experts matter and trend in the direction of national raters being more favorable. However, the fact that differences do not show up on all indicators and are not always in the direction of national favoritism cuts somewhat against in-group bias as the only mechanism.

As noted earlier, most GPIs are created in the Global North and tend to rely heavily on experts who are based there. Further analyses suggest that, if anything, the over-representation of Global North perspectives in GPIs could matter most for how positively *wealthy* countries are evaluated. Returning to the democracy scores outcome measure, Figure 3 shows how the relationship between experts' nationality and assessments varies with country wealth. 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

Table 2. The Relationship Between V-Dem Indicators and Rater Nationality. Summary of results from 71 regressions that include country and year fixed effects. "Difference" is coded using a threshold of $p < .05$.

Category	Number (%) of indicators where nationals more negative	Number (%) of indicators where nationals more positive	Number (%) of indicators with no difference
Elections	0/12 (0%)	11/12 (92%)	1/12 (8%)
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%)	12/19 (63%)	2/19 (11%)
Civil society	2/5 (40%)	2/5 (40%)	1/5 (20%)
The media	1/6 (17%)	1/6 (17%)	4/6 (67%)
Political equality	1/5 (20%)	4/5 (80%)	0/5 (0%)
Total	14/71 (20%)	43/71 (60%)	14/71 (20%)

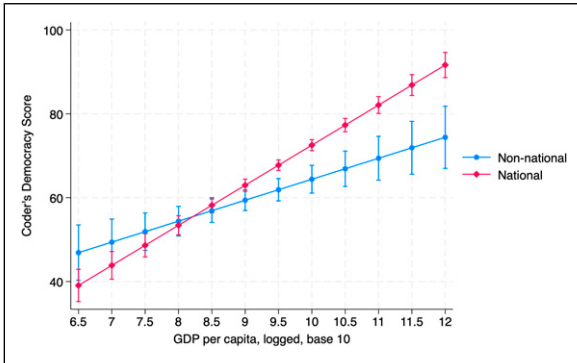


Figure 3. The Marginal Effect of Country GDP and Coder Nationality on Democracy Assessments. Democracy scores refer to the rater’s assessment of democracy in their main rated country for the year 2012 on a 0-100 scale. Nationality is coded using citizenship. The results are based on an OLS regression of the assessment on the coder’s nationality, the rated country’s GDP per capita, and their interaction with robust standard errors.

when countries are wealthy, the relationship reverses when countries are poor.²⁰ In other words, national raters in the Global North rate their countries more favorably than non-national raters. By contrast, non-national raters in the Global South rate those countries more favorably, though this difference is considerably smaller. To the extent that we can generalize from this case, the results suggest GPIs may be biased in favor of countries from the Global North given their reliance on analysts based in those countries.

Unfortunately, limitations in the V-Dem data that we can access prevent us from making inferences about the drivers of variation in national differences across countries at varying levels of economic development. 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 nevertheless draw an important conclusion: non-national raters are not more critical of most lower-income countries or African countries. Thus, our findings from the Uganda expert survey do not imply a broader tendency for national raters in Uganda or other African countries to display national favoritism. If anything, Figure 3 suggests that Uganda could be atypical of the types of countries where positive national bias is most likely to be observed globally, although 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 most V-Dem indicators. The next question is why. We find support for all three mechanisms.

First, 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 ($p < .001$), know less about Ugandan politics according to our political knowledge questions ($p < .001$), and are less likely to be certain about their responses to questions about democracy in Uganda ($p < .001$). 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 2](#) 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.

V-Dem also collects data on the sources that coders use to conduct their assessments. We compare the likelihood of using eleven different types of sources and find systematic differences in usage between nationals and non-nationals for five of them. Nationals are significantly more likely to use (1) archival documents, (2) official government data or reports, and (3) direct personal experiences, and they are significantly less likely to use (4) international news sources and (5) data and reports from international sources (see SI §5). These findings resonate with the idea that non-nationals are more likely to rely on international than domestic sources, particularly domestic government sources.

Second, we consider responses to questions about democracy in the abstract. Using anchoring vignettes developed by [King et al. \(2004\)](#), we asked the Uganda 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](#), p. 254); and freedom of speech, which is associated with more maximalist definitions, including Freedom House's.²¹ 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 scenarios in notably different ways, especially for the scenarios related to elections. For example, consider the following scenario: "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 government's policies." While the non-Ugandan experts we surveyed had a fairly strong consensus that it 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 be between "moderately free" and "very free" (with an average score of 2.5). SI §3 contains the full results of this analysis.

V-Dem also asks its coders to state the extent to which they agree that different principles are essential elements of democracy. They are the electoral, liberal, majoritarian, consensus, participatory, deliberative, and egalitarian principles of democracy. We find that nationals agree more strongly than non-nationals with three of the six principles: majoritarian, consensus, and participatory (see SI §5). There are also often differences in sympathy across regions. Perhaps importantly for our purposes, nationals in African countries agreed less strongly than non-nationals with the electoral democracy principle, which may help explain our findings with respect to the vignettes discussed above. National and non-national respondents' different assessments of the vignettes and the importance of different principles of democracy support 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.

Finally, building on the analysis in Table 2, we consider on which questions national and non-national experts provided significantly different responses in our Uganda expert survey. SI §3 lists the full text of the questions. Nationals provided higher (i.e., more positive) 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, especially on questions relating to personal autonomy and individual rights. Notably, there are no differences in assessments for the questions about electoral process or government functioning. 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.²² That we do not see national differences on more or all questions is potentially inconsistent with the in-group–out-group mechanism's prediction of across-the-board differences. At the same time, we do not see much difference in the magnitude of the nationality coefficient for the two categories—political rights and civil liberties—in Figure 2. Moreover, it may be that some of the questions, such as those on electoral process, involve less subjectivity or coder judgment and are therefore less likely to exhibit national bias (Little & Meng, 2024, p. 153).²³

Implications for Ratings' Authority

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

GPIs are more likely to have authority when they come from trusted sources (Kelley & Simmons, 2019, p. 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 is the idea that if an assessment comes from a recognized NGO, people may perceive it as high quality and legitimate (Honig & Weaver, 2019, p. 580). Supporting this logic, Nielson et al. (2019) find that civil society actors are more likely to view election observers as legitimate if they are associated with prominent organizations.

We argue, by contrast, that rater nationality is potentially significant for how ratings are used. Ratings produced by fellow nationals might be more trusted because they are perceived as more likely to reflect shared values, which is an especially important consideration when the concept being assessed is contested, as is the case with democracy GPIs (Bush, 2019). This hypothesized dynamic offers an explanation for why ratings entities are produced so overwhelmingly in North America and Western Europe (Kelley & Simmons, 2019, p. 497), which is where the greatest demand for GPIs is 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 participants included current and former staffers of aid agencies, international NGOs working on democracy and development, and others; this sampling approach reflects our understanding of the main non-academic users of democracy ratings (e.g., Bush, 2017; Büthe, 2012). We asked respondents to assess the trustworthiness of three democracy ratings: the one produced by national experts in our Uganda 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 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.²⁴

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 4, both the Freedom House score and the non-national experts' score were trusted significantly more than the Ugandan experts' score ($p < .001$ for both comparisons). By contrast, among national audiences, national raters' indicators had the most authority. Within this group, the Ugandan experts' score was trusted significantly more than the official Freedom House score ($p = .028$). It was also trusted more than the non-national experts' score, although this difference was not statistically significant ($p = .157$). Contrary to the expectation about the importance of rater branding, the official Freedom House score was not significantly more trusted than the non-branded

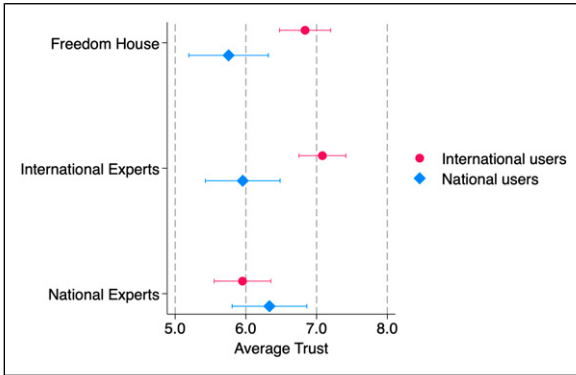


Figure 4. Trust in ratings by rater nationality. This figure shows means and 95% confidence intervals. Trust is measured on a 10-point scale.

international experts from our earlier survey among either national ($p = .441$) or non-national ($p = .226$) audiences.

An open-ended follow-up question asked respondents about their reasoning. We provide illustrative responses associated with each type of preferred rating in Table 3. The most-commonly raised concern was about the ratings’ accuracy, which was referenced by 29% of the respondents (for more details on coding, see SI §6). Non-Ugandan 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% 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 consistent with our expectations about the perceived advantages of each rating type.

Overall, these patterns suggest raters’ national identities significantly affect ratings’ authority 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. Aid officials, for example, increasingly speak about the importance of “local ownership” (Honig & Gulrajani, 2018). These dynamics have encouraged international actors to invest more in assessments with local (or peer-country) input (Swedlund, 2013, p. 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. Shaming is an important mechanism by which GPIs are thought to affect politics in the countries being

Table 3. Illustrative Responses About Preferred Democracy Score.

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 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.”

evaluated (Kelley & 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

GPIs are influential and widely-used tools in real-world decision-making and academic research. This article showed that the national identity of the experts called on to produce GPIs can matter for assessments of democracy as well as for their perceived authority.

In the case of Uganda, national experts rate the country as substantially freer than non-national experts. We 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 relative to that, non-national experts negatively biased, or both. Looking beyond the Uganda case, however, we also detect national differences among V-Dem raters. These differences are often but not always in the direction of national raters being more positive. While national differences are not present for all indicators or regions, they are present on most outcomes we analyzed.

The results suggest the need to look more carefully at the role that 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. Yet V-Dem's best practice of collecting and making available meta-data on experts' characteristics is rarely adopted by other datasets and GPIs. At the very least, transparency about the question of "who rates" is important for these indicators going forward.

What else can and should GPI creators do to address the potential for national bias? Most GPIs attempt to deal with inter-coder variance through methods such as training raters, developing better coding criteria, and reviewing draft ratings. Freedom House conducts ratings review meetings at the regional level, for example, and we may overestimate national differences by skipping this review step in our expert survey. Yet it remains difficult to ensure that two coders will evaluate a country similarly given the latent quality of many concepts relevant for GPIs, including democracy (Coppedge et al., 2019, p. 68). Moreover, it is not clear that training, guidelines, or corrections implemented by non-national GPI producers can fix biases that are introduced by relying mostly on non-national experts in the first place. Our analysis shows that the size and direction of the national rater difference varies within and across countries. Tracking raters' national identities is a good first step to identifying and dealing with this issue. Another approach is to survey many different experts and use a dynamic measurement model that can account for rater differences, similar to V-Dem. Related methods have been used to measure human rights abuses (Schnakenberg & Fariss, 2014). Of course, these approaches will not always be feasible since they require substantial resources given that they aggregate across multiple raters and ratings, not to mention requiring significant technical skills.

In addition, 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 second, that the gap between national and non-national raters is larger in wealthier countries. Systematic overrepresentation of wealthy countries among raters and 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 that rely on expert judgments, few of which make explicit commitments to include national raters in the way V-Dem has. An implication is that how we perceive democracy and other concepts in some regions may be more affected by national bias than others.

Beyond V-Dem, 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 easily rectified. For example, Swedlund (2013, p. 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 real. Nevertheless, the solution should not be to maintain the status quo but to continue working to include a wider variety of experts.

Finally, turning to GPIs’ authority, would national assessments for the United States 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 United States such as Bright Line Watch and the Authoritarian Warning Survey, which are dominated by American experts’ evaluations of American democracy, provides *prima facie* evidence that they are not. Are concerns about national 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. We see these as directions for future research that could further our understanding of how GPIs are created and the extent to which biases may feature in these influential tools in global governance.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Sarah Sunn Bush  <https://orcid.org/0000-0002-0889-5916>

Melina R. Platas  <https://orcid.org/0000-0002-5503-1029>

Data Availability Statement

Replication data and code may be found on Dataverse at [Bush and Platas \(2024\)](#).

Supplemental Material

Supplemental material for this article is available online.

Notes

1. According to our coding of [Kelley and Simmons \(2019\)](#), 64% of the GPIs in their dataset rely on expert judgments. The remainder are produced entirely on the basis of objective, public information (e.g., economic statistics).

2. For example, [Haschke and Arnon \(2020, p. 41\)](#) observe that the potential for “coder effects,” which include coder nationality effects, “has largely been ignored in the literature” on bias in human rights GPIs, even though this literature is large and growing.
3. Experts’ national identity can be thought of in terms of citizenship, residence, education, and more ([Colgan, 2019b, p. 305](#)). It is also possible to conceptualize it as a spectrum ([Cheng & Brettelle, 2019](#)).
4. This dynamic may not hold in highly-repressive societies, where nationals may have more limited access to political information.
5. In-group favoritism may extend to other countries. Studies have found that experts view their countries’ allies as more democratic than is merited ([Colgan, 2019a; Oren, 1995](#)).
6. Replication data can be found at [Bush and Platas \(2024\)](#).
7. See Supporting Information (SI) §1. SI §8 contains a discussion of research ethics.
8. At present, how to classify backsliding democracies is another especially pressing and debated issue related to democracy measurement (e.g., [Little & Meng, 2024](#)). Focusing on such a case would be a promising direction for future research along these lines.
9. As SI §2 shows, there is also substantial disagreement between Ugandan citizens and democracy GPIs. 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.
10. We interviewed five people involved in the creation or production of democracy GPIs. See SI §8.
11. For 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. \(2019, pp. 58–61\)](#).
12. E.g., [Dietrich \(2016, p. 90\)](#), [Swedlund \(2017, p. 466\)](#), and [Safarpour et al. \(2022\)](#). Additional notes on sampling are in SI §2.
13. Our 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 eliminated thirty-one respondents (26%).
14. Using the V-Dem questionnaire would have made the data from our Uganda expert survey more directly comparable to the data we analyze in the next section from V-Dem. However, since our objective is to understand how national identity shapes democracy ratings in general, demonstrating the significance of national identity across two distinct questionnaires enhances our study’s generalizability.

15. See <https://www.v-dem.net/en/news/call-country-experts-v-dem-update-2020/> (accessed July 31, 2020).
16. The substantive effect is also larger than that of gender, political ideology, and industry, among other individual-level variables, as shown in SI §3.
17. See <https://www.mcc.gov/who-we-select/indicator/civil-liberties-indicator> (accessed May 17, 2023).
18. These covariates are similar to those in Razafindrakoto and Roubaud (2010). We show in SI §3 that the results are similar, though more marginal in terms of statistical significance, when we include only demographic controls.
19. Thus, we exclude raters who evaluated multiple countries as we cannot determine their nationality.
20. We show in SI §5 the interaction between nationality and country Polity score, as well as both interactions using residency instead of citizenship to measure nationality. National experts are, if anything, more critical than non-national experts in staunchly authoritarian countries, which suggests that the effect of national identity is not driven by national experts' reliance on censored information in authoritarian environments. The gap between residents' and non-residents' scores increases with income, as well.
21. The vignettes were randomized such that they appeared either before or after the questions concerning democracy in Uganda.
22. This pattern is consistent with an in-group–out-group bias mechanism in which the non-Ugandan raters are evaluating their own countries (i.e., those that might be influencing Ugandan politics) positively.
23. But see Knutsen et al. (2024) for an alternative perspective. We also note that in Table 2, the “elections” section is not one with high levels of coder agreement.
24. In 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.

References

- Andersson, S., & Heywood, P. M. (2009). The politics of perception: Use and abuse of Transparency International's approach to measuring corruption. *Political Studies*, 57(4), 746–767. <https://doi.org/10.1111/j.1467-9248.2008.00758.x>
- Arnon, D., Haschke, P., & Park, B. (2023). The right accounting of wrongs: Examining temporal changes to human rights monitoring and reporting. *British Journal of Political Science*, 53(1), 163–182. <https://doi.org/10.1017/s0007123421000661>
- Barnett, M. N., & Finnemore, M. (2004). *Rules for the world: International organizations in global politics*. Cornell University Press.
- Besley, T. (2015). Law, regulation, and the business climate: The nature and influence of the World Bank Doing Business project. *The Journal of Economic Perspectives*, 29(3), 99–120. <https://doi.org/10.1257/jep.29.3.99>
- Bilodeau, A. (2014). Is democracy the only game in town? Tension between immigrants' democratic desires and authoritarian imprints. *Democratization*, 21(2), 359–381. <https://doi.org/10.1080/13510347.2012.712515>

- Bradley, C. G. (2015). International organizations and the production of indicators: The case of Freedom House. In S. E. Merry, K. E. Davis, & B. Kingsbury (Eds.), *The quiet power of indicators: Measuring governance, corruption, and the rule of law* (pp. 27–74). Cambridge University Press.
- Brooks, S. M., Cunha, R., & Mosley, L. (2015). Categories, creditworthiness, and contagion: How investors' shortcuts affect sovereign debt markets. *International Studies Quarterly*, 59(3), 587–601. <https://doi.org/10.1111/isqu.12173>
- Bush, S. S. (2017). The politics of rating freedom: Ideological affinity, private authority, and the Freedom in the World ratings. *Perspectives on Politics*, 15(3), 711–731. <https://doi.org/10.1017/s1537592717000925>
- Bush, S. S. (2019). National perspectives and quantitative datasets: A silver lining? *Journal of Global Security Studies*, 4(3), 372–383. <https://doi.org/10.1093/jogss/ogz022>
- Bush, S. S., & Platas, M. R. (2024). Replication data for 'national identity and democracy ratings. *Harvard Dataverse*. <https://doi.org/10.7910/DVN/1G5GK1>
- Bush, S. S., & Zetterberg, P. (2021). Gender quotas and international reputation. *American Journal of Political Science*, 65(2), 326–341. <https://doi.org/10.1111/ajps.12557>
- Büthe, T. (2012). Beyond supply and demand: A political-economic conceptual model. In K. Davis, A. Fisher, B. Kingsbury, & S. E. Merry (Eds.), *Governance by indicators: Global power through classification and rankings* (pp. 29–51). Oxford University Press.
- Cheng, C., & Brettelle, A. (2019). How cognitive frameworks shape the American approach to international relations and security studies. *Journal of Global Security Studies*, 4(3), 321–344. <https://doi.org/10.1093/jogss/ogz026>
- Clark, A. M., & Sikkink, K. (2013). Information effects and human rights data: Is the good news about increased human rights information bad news for human rights measures? *Human Rights Quarterly*, 35(3), 539–568. <https://doi.org/10.1353/hrq.2013.0046>
- Clarke, K. (2023). Which protests count? Coverage bias in Middle East event datasets. *Mediterranean Politics*, 28(2), 302–328. <https://doi.org/10.1080/13629395.2021.1957577>
- Colgan, J. D. (2019a). American bias in global security studies data. *Journal of Global Security Studies*, 4(3), 358–371. <https://doi.org/10.1093/jogss/ogz030>
- Colgan, J. D. (2019b). American perspectives and blind spots on world politics. *Journal of Global Security Studies*, 4(3), 300–309. <https://doi.org/10.1093/jogss/ogz031>
- Cooley, A. (2015). The emerging politics of international rankings and ratings: A framework for analysis. In A. Cooley, & J. Snyder (Eds.), *Ranking the world: The politics of international rankings and ratings* (pp. 1–38). Cambridge University Press.
- Coppedge, M., Gerring, J., Altman, D., Bernhard, M., Fish, S., Hicken, A., Kroenig, M., Lindberg, S. I., McMann, K., Paxton, P., Semetko, H. A., Skaaning, S.-E., Staton, J., & Teorell, J. (2011). Conceptualizing and measuring democracy: A new approach. *Perspectives on Politics*, 9(2), 247–267. <https://doi.org/10.1017/s1537592711000880>

- Coppedge, M., Gerring, J., Glynn, A. N., Henrik Knutsen, C., Lindberg, S. I., Pemstein, D., Seim, B., Skaaning, S.-E., & Jan, T. (2019). *Varieties of democracy: Measuring two centuries of political change*. Cambridge University Press.
- Cross, M. & K. D. (2013). Rethinking epistemic communities twenty years later. *Review of International Studies*, 39(1), 137–160. <https://doi.org/10.1017/s0260210512000034>
- Dietrich, S. (2016). Donor political economies and the pursuit of aid effectiveness. *International Organization*, 70(1), 65–102. <https://doi.org/10.1017/s0020818315000302>
- Engel, E. T. (2020). Letter to Secretary of State Mike Pompeo and Treasury Secretary Steven Mnuchin. *Foreign Affairs Committee of the U.S. House of Representatives*. https://foreignaffairs.house.gov/_cache/files/7/d/7d49f280-b8a5-4789-9d3a-dd8bd6310b3d/A726593A9B7DEEC19C35AD973FE84BE6.12-9-2020-tele-letter-to-pompeo-mnuchin-uganda.pdf.
- Fariss, C. J. (2014). Respect for human rights has improved over time: Modeling the changing standard of accountability. *American Political Science Review*, 108(2), 297–318. <https://doi.org/10.1017/s0003055414000070>
- Findley, M. G., Harris, A. S., Milner, H. V., & Nielson, D. L. (2017). Who controls foreign aid? Elite versus public perceptions of donor influence in aid-dependent Uganda. *International Organization*, 71(4), 633–663. <https://doi.org/10.1017/s0020818317000273>
- Fuchs, A., & Gehring, K. (2017). The home bias in sovereign ratings. *Journal of the European Economic Association*, 15(6), 1386–1423. <https://doi.org/10.1093/jeea/jvx009>
- Girod, D. M., Krasner, S. D., & Stoner-Weiss, K. (2009). Governance and foreign assistance: The imperfect translation of ideas into outcomes. In A. Magen, T. Risse, & M. A. McFaul (Eds.), *Promoting democracy and the rule of law: American and European strategies* (pp. 61–92). Palgrave Macmillan.
- Gray, J., & Hicks, R. P. (2014). Reputations, perceptions, and international economic agreements. *International Interactions*, 40(3), 325–349. <https://doi.org/10.1080/03050629.2014.899227>
- Green, J. F. (2013). *Rethinking private authority: Agents and entrepreneurs in global environmental governance*. Princeton University Press.
- Gunitsky, S. (2015). Competing measures of democracy in the former Soviet republics. In A. Cooley, & J. Snyder (Eds.), *Ranking the world: The politics of international rankings and ratings* (pp. 112–150). Cambridge University Press.
- Haas, P. M. (1992). Introduction: Epistemic communities and international policy coordination. *International Organization*, 46(1), 1–35. <https://doi.org/10.1017/s0020818300001442>
- Harmon, R., Aron, D., & Park, B. (2022). TIP for tat: Political bias in human trafficking reporting. *British Journal of Political Science*, 52(1), 445–455. <https://doi.org/10.1017/s0007123420000344>

- Haschke, P., & Arnon, D. (2020). What bias? Changing standards, information effects, and human rights measurement. *Journal of Human Rights, 19*(1), 33–45. <https://doi.org/10.1080/14754835.2019.1671179>
- Honig, D., & Gulrajani, N. (2018). Making good on donors' desire to do development differently. *Third World Quarterly, 39*(1), 68–84. <https://doi.org/10.1080/01436597.2017.1369030>
- Honig, D., & Weaver, C. (2019). A race to the top? The aid transparency index and the social power of global performance indicators. *International Organization, 73*(03), 579–610. <https://doi.org/10.1017/s0020818319000122>
- Hyde, S. D. (2011). *The pseudo-democrat's dilemma: Why election observation became an international norm*. Cornell University Press.
- Just, A. (2019). Political regimes and immigrant party preferences. *Comparative Political Studies, 52*(5), 651–686. <https://doi.org/10.1177/0010414018797938>
- Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). *The worldwide governance indicators: Methodology and analytical issues* World Bank Policy Research Working Paper No. 5430. World Bank.
- Kelley, J. G., & Simmons, B. A. (2015). Politics by number: Indicators as social pressure in international relations. *American Journal of Political Science, 59*(1), 55–70. <https://doi.org/10.1111/ajps.12119>
- Kelley, J. G., & Simmons, B. A. (2019). Introduction: The power of global performance indicators. *International Organization, 73*(03), 491–510. <https://doi.org/10.1017/s0020818319000146>
- King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review, 98*(1), 191–207. <https://doi.org/10.1017/s000305540400108x>
- Knutsen, C. H., Marquardt, K. L., Seim, B., Coppedge, M., Edgell, A., Medzihorsky, J., Pemstein, D., Teorell, J., Gerring, J., & Lindberg, S. I. (2024). Conceptual and measurement issues in assessing democratic backsliding. *PS: Political Science & Politics, 57*(2), 162–177. <https://doi.org/10.1017/S104909652300077X>
- Little, A. T., & Meng, A. (2024). Measuring democratic backsliding. *PS: Political Science & Politics, 57*(2), 149–161. <https://doi.org/10.1017/S104909652300063X>
- Lührmann, A., Tannenberg, M., & Lindberg, S. I. (2018). Regimes of the world (RoW): Opening new avenues for the comparative study of political regimes. *Politics and Governance, 6*(1), 60–77. <https://doi.org/10.17645/pag.v6i1.1214>
- Mutz, D. C., & Kim, E. (2017). The impact of in-group favoritism on trade preferences. *International Organization, 71*(4), 827–850. <https://doi.org/10.1017/s0020818317000327>
- Nielson, D. L., Hyde, S. D., & Kelley, J. G. (2019). The elusive sources of legitimacy beliefs: Civil society views of international election observers. *The Review of International Organizations, 14*(4), 685–715. <https://doi.org/10.1007/s11558-018-9331-6>

- Oren, I. (1995). The subjectivity of the “democratic” peace: Changing U.S. Perceptions of imperial Germany. *International Security*, 20(2), 147–184. <https://doi.org/10.2307/2539232>
- Razafindrakoto, M., & Roubaud, F. (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. <https://doi.org/10.1016/j.worlddev.2010.02.004>
- Safarpour, A., Bush, S. S., & Hadden, J. (2022). Participation incentives in a survey of international non-profit professionals. *Research & Politics*, 9(3), Article 205316802211257. <https://doi.org/10.1177/20531680221125723>
- Schaffer, F. C. (2000). *Democracy in translation: Understanding politics in an unfamiliar culture*. Cornell University Press.
- Schnakenberg, K. E., & Fariss, C. J. (2014). Dynamic patterns of human rights practices. *Political Science Research and Methods*, 2(1), 1–31. <https://doi.org/10.1017/psrm.2013.15>
- Seabrooke, L., & Wigan, D. (2015). Powering ideas through expertise: Professionals in global tax battles. *Journal of European Public Policy*, 23(3), 357–374. <https://doi.org/10.1080/13501763.2015.1115536>
- Sears, D. O., & Brown, C. (2013). Childhood and adult political development. In *The Oxford handbook of political psychology* (2nd ed., pp. 59–95). Oxford University Press.
- Swedlund, H. J. (2013). The domestication of Governance Assessments: Evidence from the Rwandan Joint Governance Assessment. *Conflict, Security and Development*, 13(4), 449–470. <https://doi.org/10.1080/14678802.2013.834117>
- Swedlund, H. 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. <https://doi.org/10.1080/09692290.2017.1302490>
- Terman, R. (2017). Islamophobia and media portrayals of Muslim women: A computational text analysis of US news coverage. *International Studies Quarterly*, 61(3), 489–502. <https://doi.org/10.1093/isq/sqx051>
- Weidmann, N. B. (2024). Recent events and the coding of cross-national indicators. *Comparative Political Studies*, 57(6), 921–937. <https://doi.org/10.1177/00104140231193006>

Author Biographies

Sarah Sunn Bush is associate professor of political science at the University of Pennsylvania.

Melina R. Platas is assistant professor of political science at New York University Abu Dhabi.