Received: 19 November 2019 | Revised: 16 March 2020 | Accepted: 27 April 2020

DOI: 10.1111/gove.12504

ORIGINAL ARTICLE

Governance WILEY

How global performance assessments shape attitudes toward government decision-making: Survey experimental evidence

Emmerich Davies[1] | **Thomas Gift**[2] | **Carlos X. Lastra-Anadón**[3]

[1]Harvard University

[2]University College London

[3]IE University

Abstract

Global Performance Assessments (GPAs), which rank countries on a range of policy areas, can encourage domestic demands for policy reform. Yet can they also affect at what level of government—local or national—citizens want reform to take place? We theorize that, by emphasizing how countries fare relative to others, GPAs prompt citizens to view domestic policy underperformance as a "national problem requiring national solutions." This increases calls for vesting policymaking authority in the hands of central governments. We argue that this effect should be most salient when underperformance is presented as a threat to a country's security because it induces citizens to "rally 'round the flag." To test our theory, we field an original survey experiment in the United States using fictitious news articles manipulating both the source of performance monitoring information and how it is presented. In line with our prediction, respondents are most likely to demand policy centralization when underperformance is framed using GPAs and citizens are primed to think of low scores as a threat to their country's security. These results indicate that GPAs could eventually increase calls for expanding the purview of national-level politicians over policymaking.

1 | INTRODUCTION

In recent decades, Global Performance Assessments (GPAs) have become one of the most important tools that the international community has leveraged to increase accountability over public service delivery in countries across the world.[1] By assigning scores and ranking nations from top to bottom on the perceived quality of their policymaking, GPAs strive to provide an objective benchmark of which governments deliver effective public services to their citizens (and which lag behind). In the process, GPAs are not only thought to encourage governments to improve policy outcomes in response to pressure from international organizations and other actors, but also from domestic constituencies.

Yet while evidence indicates that GPAs can promote policy reform by galvanizing domestic preferences for change (Bisbee, Hollyer, Rosendorff, & Vreeland, 2019; Kelley, 2017; Kelley & Simmons, 2015; Papanicolas & Jha, 2017), an overlooked question is which politicians—national or local—citizens want to implement these reforms. Countries vary markedly in how decentralized policymaking is (Rodden, 2004; Wibbels, 2006), with some vesting more authority in subnational governments and others in national governments. Politicians who want to enact policy reforms not only require that voters demand change, but that they are open to change at the level of government where they operate.

Consider, for instance, the Programme for International Student Assessment (PISA), the most well-known educational GPA attempting to measure the performance of school systems across the world. When the results of this International Large-Scale Assessment (ILSA) were released in 2001, it prompted widespread calls for school reform in many countries, including the United States, by highlighting failures in educational delivery. Yet given extensive variation in which levels of government control education in nations around the globe, the capacity of national and subnational politicians to implement school reforms likely differed depending on existing institutional arrangements.

Can GPAs themselves affect whether citizens want policymaking to occur at a centralized or decentralized level? In this article, we argue that, by framing policy underperformance through GPAs, voters should be more likely to call for policy centralization because it prompts them to view underperformance as a "national problem requiring national solutions." When country comparisons are made explicit, people should be more apt to assume nation-centric attitudes that lead them to want to resolve challenges through the central government. This contrasts with national performance assessments (NPAs), where county comparisons are not provided, and citizens should favor more localized policymaking.

We further argue that preferences over which level of government controls policymaking should depend on how assessments are framed to the public. When underperformance is presented as a threat to a country's security, citizens should be more inclined to call for policy centralization because it prompts them to "rally 'round the flag." The logic is that, when citizens are scared amid crises of confidence, they band together as a nation. This contrasts with the other primary way in which policy outcomes are often presented, in which citizens are made to feel ashamed of poor national performance. In this case, citizens should feel less patriotic and be more disinclined to vest policymaking in national authorities.

To test our theory, we conducted a national survey experiment in the United States that randomly exposed a diverse convenience sample of respondents to different aspects of performance monitoring in K-12 education and then measured preferences for policy centralization. We looked at education because many countries should prioritize and expect to do well in this area and because falling behind in education can pose threats to a country's well-being. Additionally,

disputes over which level of government should have responsibility for education policy are salient in many nations. GPAs in education are also increasingly publicized globally, with tests like PISA gaining significant media attention.

Besides underperforming on international exams and having a history of controversy over where education policy should be vested, the United States is a particularly useful case for our experiment because it provides a "hard test" for our predictions. Because U.S. education has historically been a state and local issue, and because poll after poll shows that Americans are predisposed to favor their local schools, they should be broadly disinclined to call for centralized policymaking over education. To the extent that our experiment can induce preferences for centralization even in this context, our results are likely to generalize to other settings and policy areas where favorability toward the status quo is weaker.

We primed respondents by randomly manipulating both the source of the monitoring information and how it was presented. We then measured preferences for policy centralization versus decentralization over education. Regarding the source of information, treatment groups received information on educational underperformance in the United States from PISA, the most widely-publicized global standardized test in education. Control groups received information from the National Assessment of Educational Progress (NAEP), the main domestic standardized test in the United States. This directly compared citizen responses to GPAs versus NPAs.

For both the GPA treatment and the NPA control, we also randomly varied how educational underperformance was presented in ways that simulated how people consume monitoring information in real life. Treatment groups were exposed to a "scare" vignette, which suggested that citizens in the United States should be frightened about its poor performance given that education affects its ability to compete globally. Control groups were exposed to a "shame" vignette, which indicated that citizens in the United States should be ashamed of its poor performance given its vast wealth and resources. This directly tested citizen responses to different presentations of performance monitoring information.

As expected, respondents who received both the GPA and scare treatment supported policy centralization over education the most. Conversely, respondents who received both an NPA and a shame control supported policy centralization over education the least. Respondents were moderately, but not significantly, more likely to call for policy centralization when receiving the combination of an NPA control and a scare treatment and the combination of a GPA treatment and a shame control. Our primes did not significantly affect other policy attitudes over education—namely, the desire for school reform, assessment of the perceived quality of public schools, or support for increased school funding.

To our knowledge, we are the first to test how performance monitoring shapes citizen attitudes toward policy centralization. In doing so, we shift the analytical lens on GPAs from a focus on demands for policy reforms to how public provisions are delivered at different levels of government. Our study complements sizable scholarships on preferences for decentralization (Bardhan & Mookherjee, 2006; Falleti, 2010; Manor & Crook, 1998) and how international variables shape domestic-level politics (Gourevitch, 1978; Putnam, 1988). We also build on an emerging literature in comparative education probing the impact of ILSAs on policy attitudes (Heyneman & Lee, 2014; Pizmony-Levy, 2018; Volante, 2015).

Practically, our results matter because they suggest that GPAs can not only spur domestic demand for policy reforms, but they can also alter where citizens want those reforms to occur. Evaluating policy underperformance in the context of GPAs presumably provides political cover for national-level politicians to implement reforms more than for state, provincial, and local

politicians. In the short run, GPAs could increase the leverage of national-level politicians to influence decision-making over relevant public policies. In the long run, they could even prompt calls for revising institutional arrangements to condense more authority for policymaking in the hands of national governments.

2 | MONITORING POLICY OUTCOMES

In recent years, GPAs have increased substantially in their usage and visibility (Cooley & Snyder, 2015; Davis, 2012; Green Saraisky, 2015; Kelley, 2017; Kelley & Simmons, 2015; Merry, Davis, & Kingsbury, 2015; Singer & Braun, 2018). Indicators such as The World Bank's "Ease of Doing Business Index," Transparency International's "Corruption Perceptions Index," the Fund for Peace's "Fragile States Index," and the UN's "Sustainable Development Goals" are just some high-profile examples. Today, as many as 289 GPAs now exist across a panoply of policy domains, including education, the environment, the economy, energy, infrastructure, and health (Kelley & Simmons, 2019).

An expanding literature suggests that GPAs can trigger demands for policy change globally (Bieber & Martens, 2011; Kelley, 2017; Kelley & Simmons, 2015; Kelley & Simmons, 2019; Kijima & Lipscy, 2017). By facilitating policy borrowing and convergence (Bieber & Martens, 2011; Davis, Wilson, & Dalton, 2018; Kijima & Lipscy, 2017; Steiner-Khamsi, 2003; Volante, 2015), by calling out high- and low-performing countries on the international stage (Kelley, 2017; Kelley & Simmons, 2015), and by erecting benchmarks against which to measure progress (Papanicolas & Jha, 2017; Steiner-Khamsi, 2003), GPAs can encourage governments to reform domestic policy (Kelley, 2017).

At the international level, much of the impetus pushing countries to undertake policy reforms occurs via "carrots" and "sticks." In terms of carrots, countries that make demonstrable progress in certain policy areas may be more likely, for example, to qualify for foreign assistance from international organizations and to receive other forms of awards and recognition. In terms of sticks, countries that do not perform well or that experience backsliding in policy performance may be less likely to receive such benefits. In both ways, countries are incentivized to identify and implement best practices in order to improve their GPA rankings and earn credibility from the international community.

Although much existing literature focuses on how GPAs promote policy reform by making politicians mindful of foreign audiences—especially by encouraging politicians to care how they are viewed by global actors—at least as important is how GPAs affect public opinion within countries. Domestic audiences likely matter even more for politicians because it is ultimately these citizens who hold them accountable. Mounting evidence suggests that GPAs cause citizens to recognize failures in policymaking and to shape preferences over what needs to be done to improve (Bieber & Martens, 2011; Doshi, Kelley, & Simmons, 2019; Volante, 2015). When countries fare poorly on global rankings, citizens may pressure governments to change.

Yet despite the fact that GPAs are increasingly seen as a mechanism to promote domestic demands for policy reform, one question that remains unanswered is whether GPAs influence at what level of government—national or subnational—citizens want reforms to take place.[2] This question is crucial because, to implement reforms, politicians not only need citizens to support changes, but they must also have the authority to implement those changes. If citizens want reform, but enacted at the national level, subnational politicians will have less room to

pursue reform. By comparison, if citizens want change, but enacted at the subnational level, national politicians will have less room to pursue reform.

Below, we theorize how different ways of framing performance assessments can shape public attitudes toward the centralization or decentralization of public services. The most obvious comparison is how GPAs differ from NPAs, which monitor policy outcomes only within individual countries. We claim that, by expressly comparing countries to one another, GPAs should encourage preferences for policy centralization by prompting voters to see underperformance as a "national problem requiring national solutions." This contrasts with NPAs, which do not invoke country comparisons and should lead citizens to prefer more policy decentralization.

Additionally, we claim that how intermediaries like the media (Kelley, 2017; Pizmony-Levy & Bjorklund, 2018; Pizmony-Levy & Torney-Purta, 2018) present information from GPAs and NPAs also matters. We suggest that two of the most common ways that poor outcomes are presented is through "scare" and "shame" framing. We argue that, when actors attempt to make voters scared by presenting policy underperformance as a threat to their country's security, citizens should "rally 'round the flag" and push for policy centralization. By comparison, when actors attempt to make voters ashamed by presenting poor outcomes as a national embarrassment, citizens should feel less patriotic and prefer policy decentralization.

Our conception of framing builds on a considerable literature in political psychology, where individuals use heuristics to process information (Tversky & Kahnemann, 1974). Because voters find it difficult to process all relevant information, they rely on cognitive shortcuts when faced with complex problems (Healy & Malhotra, 2013). We claim that when voters learn about policy underperformance, they assume a certain level of government to hold accountable. Framing over the source (GPA vs. NPA) and presentation (scare vs. shame) of performance information causes individuals to update their preferences over which level of government should be held responsible for policymaking.

3 | SOURCE OF INFORMATION: GPA VERSUS NPA

First, we contend that whether information about policy underperformance is framed via GPAs or NPAs can influence voter opinions about which level of government should oversee reforms. In particular, it may affect preferences for how centralized (or decentralized) public service delivery should be. In our formulation, whether performance assessments cause people to want national or subnational governments to control policymaking hinges on whether they view their country predominantly as a single indivisible unit competing with other nations. Compared to NPAs, we claim that GPAs should prompt more of this orientation and lead citizens to want authority vested in national-level politicians.

The reason is that when cross-national comparisons are made explicit as they are with GPAs, people should think more about their country gaining or losing relative to other countries. This should instill a nation-centric perspective whereby people think about policy on a national scale and feel a patriotic duty to ensure that their country succeeds. When the emphasis is on the country as a whole doing better relative to other countries, citizens should be more apt to believe that "national problems require national solutions" (Fusarelli & Fusarelli, 2008).[3] Because the country will compete more effectively if the entire nation improves, citizens may believe this requires coordination by a central government.[4]

By comparison, we argue that when country comparisons are not made explicit as is the case with NPAs, citizens should be less inclined to view their country as competing with other

nations. On the contrary, they should be more likely to think about their own area or region as part of a collection of units that collectively comprise the country. Because there is not a sense of the nation competing against other nations, citizens may see less need for policy coordination at the country level. Instead, their preference may be that reform occurs through less nationalized channels that prioritize autonomy by subnational decision-makers. This induces relative support for decentralization.

Take, for example, the case of education in the United States, which both administers its own NPA, the National Assessment of Educational Progress (NAEP), and participates in a GPA, the Programme for PISA. When PISA results are released, commentators often invoke comparisons to other countries, noting, for instance, that the United States does not perform well relative to what might be expected. The implication is that, to improve its PISA ranking, the nation as a whole has to enact reforms. A 2016 *New York Times* article on PISA entitled, "What America Can Learn From Smart Schools in Other Countries" highlights this national approach.[5]

Compare this reaction to the results from NAEP, which tend to yield calls for more localized school reform. Although NAEP makes country-level data on educational performance available, much of the conversation surrounding its release concerns achievements by individual states. Even in Massachusetts, a perennial top performer in education that sets the standard in math and reading, there is reflection about how to maintain high scores. For example, in 2018, the state's secretary of education called for Massachusetts to "redouble our efforts on educational policy and education reform." He said, "We need to do better... We can't afford to stay in place, even though we are at the top of the ratings."[6]

The central distinction in how GPAs and NPAs condition attitudes toward policy centralization versus decentralization is in how they anchor people to think about the nature of policy challenges and who should address them. By drawing boundaries around countries and providing cross-national comparisons, GPAs prime individuals to see problems—and solutions—as predominantly national. Because other countries constitute competition, citizens may want their nation to compete by setting centralized policies. NPAs, however, do not offer cross-country comparisons, so their effect may be different. Instead, citizens should view problems as less national in scale, leading to greater calls for policy decentralization.

4 | PRESENTATION OF INFORMATION: SCARE VERSUS SHAME

Whether through GPAs or NPAs, we suggest that there are two common ways that the media and other actors frame the presentation of monitoring outcomes. The first uses what we call "scare" language, where poor performance is said to endanger a country's security. The objective is to activate fears about what will happen if a population does not improve in some policy area that is essential to its safety and prosperity. Global politics is often thought to be zero-sum, where societies need to be self-preserving to shield themselves against damage. When a population lags in a policy area that could destabilize it—or even pose an existential threat to the country—citizens should demand policy action.[7]

"Shame" language is the second common way that performance assessments are presented to the public. The point is to draw a contrast between expectation and reality in terms of what the global hierarchy looks like. Shame may result from any number of factors. High-income

countries may be more likely to feel shame given the expectation that wealth should lead to better policy outcomes. Furthermore, more spending on a policy (e.g., as a percentage of GDP or total state expenditures) may heighten shame in light of expectations that these investments should increase performance. Cultural factors might also play a role insofar as some countries take special pride in certain policy areas.[8]

We argue that scare and shame language may activate different preferences among citizens over where responsibilities for policymaking should lie.[9] Scare language should be more likely to result in calls for policy centralization by inducing a "rally 'round the flag" effect (Mueller, 1970; Norpoth, 1984). Just like in times of war, people should be more likely to put aside their differences during crises. When citizens are faced with evidence that their country underperforms—and that this credibly endangers economic, military, or other forms of security—people should adopt a posture of national cohesion. This encourages people to favor political authority being vested in the hands of the central government.

The converse should be true when citizens are presented with shame language. In this scenario, people may be less prone to feelings of nationalistic fervor, since their security is not on the line. Citizens may feel less pride or allegiance to their country and therefore possess a weaker preference that policy autonomy be granted to the national government. The idea is that, even if the current system does not work, there is no reason to band together as a country and to hand over responsibilities to national politicians to direct public service delivery. Instead of harboring a sense of national cohesion, citizens may find it appropriate to defer policymaking to local or state government officials.

For real-world examples of how performance assessments are presented, consider again the case of education. Regarding shame language, a 2016 piece in *Business Insider* reported that, on a recent PISA exam, "[T]he United States did poorly compared to other countries and territories, outranked by 38 countries in math, 24 in science, and 22 in reading."[10] To reinforce this point, the article referred to an expert noting that "[t]he results are especially stark when looking at US student achievement compared to much poorer countries." The story even pointed out that "[t]he poorest 10% of Vietnamese students performed better on an international exam than the average American teen."[11]

For scare language, by contrast, consider the following 2011 excerpt from the *New York Times*. It similarly documented that the United States rated disappointingly on PISA compared to other countries: "An international study published last month looked at how students in 65 countries performed in math, science and reading... We [the United States] came in 15th in reading, 23rd in science and 31st in math."[12] It then went on to explain why this might imperil the United States relative to a key geopolitical rival: "Americans think of China's strategic challenge in terms of, say, the new Chinese stealth fighter aircraft. But the real challenge is the rise of China's education system."[13]

When presented with scare language, the response from citizens is more likely to be that the country as a whole has to do better. If poor performance poses a danger for the population, resorting to improvements at the local level will not suffice for staving off political, economic, and military challenges. To the extent that citizens agree that the best response is to improve the country as a whole—rather than simply undertaking reforms in individual states, localities, or districts—scare framing should invoke stronger calls for policy centralization. By comparison, with shame language, citizens may be more inclined to believe that poor performance can be addressed by improving local circumstances.

4.1 | Main predictions

The above discussion yields our main predictions, summarized below and outlined in Table 1:

Main predictions: *The combination of GPA and scare framing should lead to the strongest preferences for policy centralization. The combination of NPA and shame framing should lead to the weakest preferences for policy centralization.*

5 | THE SURVEY

To test our predictions, we fielded a survey experiment in the United States that manipulated both the source of performance monitoring information to which individuals were exposed (GPA vs. NPA) and how this information was presented (scare vs. shame language), and then gauged preferences for policy centralization. Specifically, we supplied information as a vignette that resembled a newspaper article on the release of education performance assessments, with a headline, descriptive text, and a quote from an expert. The goal was to elicit public opinion in a context that mirrored how people would consume performance monitoring information in real life.

5.1 | Issue area: Education

We leveraged education as an emblematic policy area for several reasons. First, many countries—particularly rich ones, those that expend considerable resources on education, and those that take special pride in education—should expect to do well on educational assessments. This should apply to most advanced, postindustrialized countries, but also extend beyond these cases (Heyneman & Lee, 2014; Lockheed, Prokic-Bruer, & Shadrova, 2015). Additionally, education is a policy realm in which falling behind can frequently pose both military and economic threats to a country. As already noted, this means that education results can be presented using both scare and shame language.

We also focus on education because it is a policy area in which calls for shifting power between different tiers of government are frequent (Channa, 2015; Fiske, 1996; Gershberg, González, & Meade, 2012; Guerra & Lastra-Anadón, 2019). Many countries exhibit sharp intragovernmental competition over where responsibility for education should be vested for administration, financing, and regulation. Some experts argue that bringing education "closer to the people" helps to ensure that education systems are sensitive to the needs of local constituents (Tiebout, 1956), whereas others argue that centralized setups can capitalize on economies of scale and erase educational disparities (Lassiter, 2007).

Finally, we look at education because International Large-Scale Assessments (ILSAs), as GPAs are often referred to in education, are highly publicized (Breakspear, 2012; Green Saraisky, 2015; Grek, 2009; Ramirez, Schofer, & Meyer, 2018; Singer & Braun, 2018; Stack, 2007).[14] As such, we

TABLE 1 Preferences for centralized control over policymaking

	Global performance assessment	National performance assessment
Scare framing	Strongest	?
Shame framing	?	Weakest

might expect many voters to be aware of GPAs in education and to use them to inform their political preferences. Although several tests compare student performance globally,[15] PISA is the most prominent. First administered in 2000 and run by the OECD, PISA is given to a national sample of 15-year olds every 3 years in more than 70 economies.

Despite the increasing prominence of educational GPAs both in the United States and elsewhere, little is known about how they affect public opinion and voter preferences for policy reforms. One recent study on elite opinion, for example, notes that "[w]hile the proliferation and increasing visibility of CNAs [cross-national assessments] is unmistakable, very little scholarship has systematically examined how CNAs affect policy outcomes" (Kijima & Lipscy, 2017). Discussing voter preferences more explicitly, another study observes that "scholars have paid little attention to the links between ILSA [international large-scale assessments] and public opinion" (Pizmony-Levy, Doan, Carmona, & Kessler, 2017).

5.2 | Test case: United States

We fielded our experiment in the United States in large part because it underperforms on PISA. In 2018, for example, the United States scored just above the OECD average in reading, and just below the OECD average in math (Schleicher, 2019). It is reasonable to assume, however, that citizens would expect the United States to perform at or near the top of PISA. Not only is it a wealthy country that spends extensively on K-12 education, but for many decades, the United States took pride in leading the world in school attainment and in the percentage of students who pursued tertiary schooling (Goldin & Katz, 2010). Recently, however, many countries have caught up to—and surpassed—the United States on various educational metrics.

The United States also has a long history of controversy over who should control education (Tampio, 2017). Traditionally, education in the United States has been an overwhelmingly state and local issue. Even now, state and local revenue account for 92% of revenue for public K-12 schools, with the balance being provided by federal funds (National Center for Education Statistics, 2018, table 235.10). Yet in the last two decades, the U.S. government has increasingly expanded its reach over education through national accountability regimes, funding formulas, and standards. Both Republicans and Democrats have advanced recent efforts to concentrate more power over education in Washington (Sides, Tesler, & Vavreck, 2019).[16]

Lastly, U.S. education should be a "hard test" for our theory. A consistent finding from polling on U.S. education is that, even though Americans express serious concerns about the quality of the country's schools, they tend to have favorable attitudes toward their own local schools (Henderson, Houston, Peterson, & West, 2020).[17] This should make voters generally resistant to concentrating more power over education in the hands of the federal government. To the degree that voters do favor more policy centralization in response to our treatments, we might expect these results to be at least as significant in other contexts and policy areas where voters have less strong attachments to the status quo.

5.3 | The sample

We use original survey data from the Harvard Digital Lab for the Social Sciences (DLABSS) (http://dlabss.harvard.edu/) collected between July 5 and November, 13, 2018.[18] DLABSS is a widely-used and established online survey platform comprised of a pool of individuals 18 years

of age and older who voluntarily complete questionnaires related to the social sciences. This is a convenience sample of volunteers recruited through connections (such as newsletters and ads) with the university that are exposed to questionnaires weekly in which they may want to participate, with no commitment. The DLABSS pool is diverse in terms of demographics—for example, race, sex, and age—and approximates common online survey platforms like MTurk. One key benefit of the sample is that because all participants in DLABSS are volunteers, they may be more inherently motivated than paid respondents to answer questions truthfully and completely (Strange et al., 2019). Relative to nationally-representative panels such as the Current Population Survey, DLABSS respondents tend to have more schooling, earn less income, and be more politically active. Scholars have replicated many social science experiments using DLABSS data, and multiple academic papers rely on DLABSS (for a list of DLABSS research, see: http://dlabss.harvard.edu/results; Carney & Enos, 2017). For sample summary statistics and their comparison with U.S. population averages, see Table S1. All together, 602 people completed our survey out of a total pool of 9,887 individuals who were invited to participate by email. After removing participants who did not declare residency in the United States, our final sample size was 555.

5.4 | Survey instrument

Supporting Information 1 displays the language used in each of our primes. All of the primes followed the same basic format. To begin, respondents were told that the United States performs poorly in math and reading. Next, this point was documented with data from an exam. Finally, an education expert was quoted elaborating on what this means for the country. Respondents were randomly assigned to one of four groups, which varied both the source of information detailing policy underperformance and how it was presented. In terms of the source, we provided information from either a GPA or an NPA. In terms of presentation, we provided information using either "scare" or "shame" language.

We supplied evidence on low student performance based on one of two standardized exams. For the GPA treatment, evidence was attributed to the international test, PISA. For the NPA control, evidence was attributed to the U.S. domestic test, the NAEP. In addition to varying the source of information, the GPA treatment also differed from the NPA control in two ways. First, it noted explicitly where the United States ranked globally on math and reading and compared it to particular countries. Second, the expert articulated that the GPA provides incontrovertible evidence of how poorly the U.S. education system fares relative to other countries.[19]

We also manipulated how information was presented to respondents. For the scare treatment, the language underscored that the United States should be scared by its poor performance given the implications for competing in the economic and military realms. For the shame control, the language highlighted that the United States should be ashamed of its poor performance given its significant wealth. For the scare treatment, the country comparison was to two geopolitical rivals: China and Russia, which are widely seen as the two greatest threats to U.S. security at the moment. For the shame control, the comparison was to two poor countries: Peru and Vietnam, which offer some degree of geographic variation.

To ensure a clean comparison between PISA and NAEP, and to reinforce the scare and shame dimensions of our primes, some of the precise information included in our vignettes was fictitious.[20] We told respondents that both PISA and the NAEP measure student performance of 15 year-olds (this is true of PISA, but NAEP tests fourth and eighth graders). We also told

respondents that just a third of students are deemed proficient on both PISA and the NAEP (this number is closest to what is given by the NAEP). Country rankings were also fictitious, as was the education expert and her quote. Our primes mimicked how performance on these tests are disseminated in the real world and should be credible to respondents.

5.5 | Dependent variable

Our outcome variable is attitudes toward policy centralization (or decentralization) in public K-12 education. To capture these preferences, we asked respondents the following question: "In your view, which level of government in the United States should assume the largest responsibility for educating students in public K-12 schools?" Respondents could answer either: "Federal government," "State government," or "Local government." Because we are interested in whether our primes induce preferences for national policymaking, we dichotomized the dependent variable and coded calls for federal responsibility as 1 and calls for state and local responsibility as 0. It is possible that respondents have different preferences over which level of government should control different aspects of public K-12 education (e.g., funding vs. accountability vs. teacher licensure). We measure preferences for policy centralization generically, however, because it seems unlikely that most citizens would have such nuanced views on these specifics.

6 | EMPIRICAL FINDINGS

Before testing our main predictions, we examine whether receiving a GPA or a scare treatment on its own increases preferences for policy centralization over education. To do so, we estimate the following linear probability models:

$$\mathrm{Y}_i = \beta \mathrm{GPA}_i + \mathrm{X}_i + \mathrm{E}_i \quad (1)$$

$$\mathrm{Y}_i = \beta \mathrm{Scare}_i + \mathrm{X}_i + \mathrm{E}_i \quad (2)$$

GP A in Equation (1) denotes that a respondent received the GPA treatment, regardless of whether it is paired with a shame or a scare prime (the omitted category is respondents receiving any of the NPA controls). *Scare* in Equation (2) denotes that a respondent received the scare treatment, regardless of whether it is paired with a GPA or an NPA prime (the omitted category is respondents receiving any of the shame controls). For robustness, we estimate regressions both without and with a vector of individual-level covariates, X_i.[21]

Models 1 and 2 of Table 2 report these results, first without individual-level covariates. In Model 1, we find that the GPA treatment significantly increases the probability of respondents supporting federal control over education. The effect is about 7% points above a 20% baseline, or a sizable 35% increase over the NPA control. In Model 2, we find that, although directionally the point estimate of the scare treatment is positive, it does not significantly increase support for federal control over education compared to the shame control. These results are depicted graphically in Figure 1. The findings are substantively similar when we include individual-level covariates in the regressions (Models 3 and 4).

TABLE 2 Effects of GPA and scare primes on preferences for federal control

	(1)	(2)	(3)	(4)
	Federal control	**Federal control**	**Federal control**	**Federal control**
GPA	.0717*		.0813*	
	(.0360)		(.0387)	
Scare		.0472		.0237
		(.0359)		(.0385)
Observations	555	555	408	408
R^2	.007	.003	.128	.119
Individual covariates			X	X

Note: Linear probability models. Robust standard errors in parentheses.
$+p < .10$, $*p < .05$, $**p < .01$, $***p < .001$.

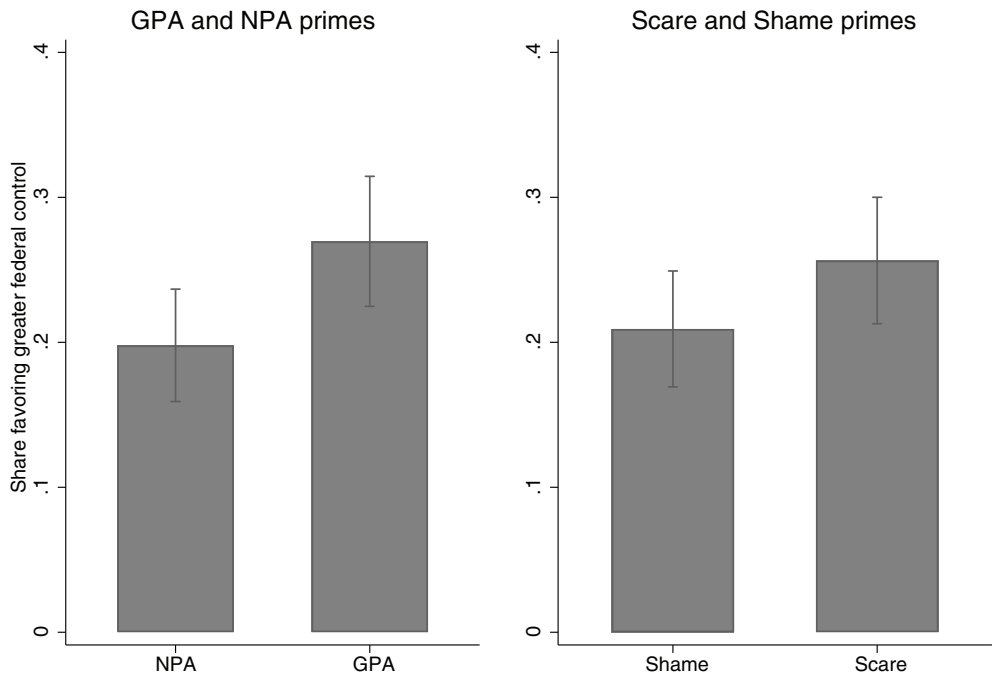

FIGURE 1 Average Level of Support for Federal Control

6.1 | Test of main predictions

We now turn to our main hypothesis: whether the combination of a GPA-scare treatment leads to the strongest preferences for policy centralization, whereas the combination of an NPA-shame control leads to the weakest. To execute this test, we estimate a linear probability model that includes two-way interactions between *GP A* and *Scare*, *N P A* and *Scare*, and *GP A* and *Shame* (respondents receiving the combination of an *N P A* and *Shame* control are the omitted category). This can be expressed as:

$$Y_i = \delta GPA_i \times Scare_i + \zeta NPA \times Scare_i + \eta GPA \times Shame_i + X_i + E_i \quad (3)$$

Model 1 of Table 3 presents these results, again first without individual-level covariates. As expected, the coefficient on *GP A × Scare* is positive, statistically significant, and—in absolute terms—larger than that on any other interaction term. Moreover, the coefficients on all of the other interaction terms are positive, suggesting that directionally *Shame × N P A* leads to the weakest preferences for policy centralization.[22] Figure 2 charts these results. The average level of support for federal control over education among respondents who received the GPA-Scare treatment is 30%. This is followed by GPA-Shame (25%), NPA-Scare (22%) and, finally, NPA-Shame (18%). Introducing individual-level covariates into the regression does not substantively alter the results (Model 2).[23]

6.2 | Subgroup analyses

We also investigate whether our primes induce larger effects on certain subgroups that we might expect to have more elastic preferences over where authority for education lies.

We first look at respondents with an interest in education—namely, those who cite education as one of the top two most important policy problems (MIPs) facing the country today. Because these respondents might be more attuned to education, their preferences could be more sensitive to priming. We find a positive and statistically significant coefficient on *GP A × Scare × Ed is priority* in Model 1 of Table 4. When receiving the GPA-Scare treatment, respondents who say that education is a top two MIP are 15% points more likely to favor federal control over education than those who do not list education as a top two MIP.

Next, we investigate whether parents specifically are also more responsive to our primes. Parents might be especially open to having their minds changed about what is the appropriate locus of decision-making over education because they have a direct stake in it. As denoted by the positive and statistically significant coefficient on *GP A × Scare × Parent* in Model 2 of Table 4, we find corroboration of this effect. When assigned to the GPA-Scare treatment,

TABLE 3 Effects of GPA and Scare primes on preferences for federal control, with interactions

	(1)	(2)
	Federal control	**Federal control**
GPA × Scare	.122*	.112^{+}
	(.0516)	(.0571)
NPA × Scare	.0459	.0115
	(.0470)	(.0514)
GPA × Shame	.0698	.0657
	(.0485)	(.0525)
Observations	555	408
R^2	.011	.129
Individual covariates		X

Note: Linear probability models. Each cell shows the marginal effect of receiving the prime pair over the omitted category (NPA × Shame). Robust standard errors in parentheses.
$+p < .10$, $*p < .05$, $**p < .01$, $***p < .001$.

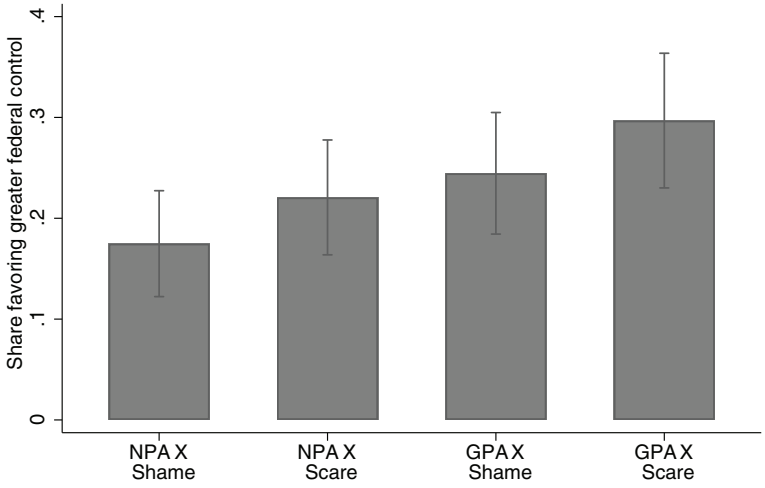

FIGURE 2 Average Level of Support for Federal Control, with Interactions

TABLE 4 Effects of GPA and Scare primes on preferences for federal control, with interactions by subgroups

	(1)	(2)	(3)
	Federal control	**Federal control**	**Federal control**
GPA × Scare × Ed is priority	.132^{+}		
	(.0768)		
GPA × Scare × Parent		.248**	
		(.0779)	
GPA × Scare × Left			.380***
			(.0679)
Ed is priority	.0197		
	(.0404)		
Parent		.115^{+}	
		(.0607)	
Left			.258***
			(.0395)
GPA × Shame	.0707	.0781	.0721
	(.0488)	(.0487)	(.0454)
GPA × Scare	.130*	.0478	.0720
	(.0633)	(.124)	(.0537)
NPA × Scare	.0458	.0545	.0297
	(.0471)	(.0475)	(.0457)
Observations	555	555	555
R^2	.011	.018	.112

Note: Linear probability models. Models do not include individual covariates. Robust standard errors in parentheses. $+p < .10$, $*p < .05$, $**p < .01$, $***p < .001$.

respondents with children are 36% points more likely to respond by calling for national control over policymaking than respondents without children.

Finally, we examine the effect of political leaning. Left-leaners, for example, could be more open to shifting educational authority to the federal government given that they are generally less distrustful of centralized policymaking. This could be the case above and beyond a pre-existing preference for the federalization of education. Results confirm this outcome. Model 3 of Table 4 yields a positive and statistically significant coefficient on *GP A* × *Scare* × *Left*. When assigned to the GPA-Scare treatment, the odds that self-identified liberals support federal control over education is 64% points higher than among conservatives.

6.3 | Placebo tests

We also attempt to rule out alternative explanations for our findings. One concern might be that the centralizing effects of GPAs that we discover are a function of some idiosyncratic feature of the particular GPA we used, rather than the fact that the performance assessment itself is global in nature. For example, perhaps PISA has more name recognition than the NAEP or citizens view it as more authoritative because they sense that it is higher quality or less subject to political manipulation. If this were the case, we would expect this feature to show up in PISA also having a significant effect on other types of attitudes bearing on education, not just those related to policy centralization.

Relatedly, one might be concerned that the "scare" treatment is somehow an inherently stronger prime than the "shame" control insofar as it better alerts people to the extent of educational underperformance in the United States. In particular, the scare treatment could lead to greater preferences for policy centralization not because of a "rally 'round the flag" effect, but instead because it triggers greater consternation about the perceived poor quality of schools. If this were the case, we would expect this also to show up in the scare treatment leading citizens

TABLE 5 Effects of GPA and scare primes on alternative outcomes

	(1) Support for increased funding	(2) Grade given to schools (1–5)	(3) Support for education reform
GPA × Scare	.0267	−.0074	.0767
	(.0581)	(.0983)	(.0656)
NPA × Scare	−.0602	−.1173	.0477
	(.0585)	(.1006)	(.0653)
GPA × Shame	.0000	−.1222	.0680
	(.0582)	(.1033)	(.0640)
Observations	527	554	471
R^2	.005	.005	.004

Note: Linear probability models in Models 1 and 3 (support binarized), and linear model with 1–5 outcome in Model 2, with 5 being the highest grade. Models do not include individual covariates. Robust standard errors in parentheses.

$+p < .10$, $*p < .05$, $**p < .01$, $***p < .001$.

to express more concern about other outcomes related to education, in addition to those bearing on policy centralization.

To check these possibilities, we estimate analogous models to those in Table 2 for the following DVs: support for increased school funding (binarized), assessment of the quality of schools nationally (scaled 1 (F)—5 (A)), and support for education reform (binarized).[24] In Models 1–3 of Table 5, we test the effect of the four different prime combinations. For none of the DVs do the primes have a statistically significant effect relative to the NPA-Shame control. The strong impact of the GPA-Scare treatment is confined to preferences over centralizing education. This increases confidence that it is the global aspect of GPAs and the "rally 'round the flag" dynamic that boosts preferences for policy centralization.

7 | CONCLUSION

Despite mounting evidence that GPAs can promote domestic demands for policy reform by alerting citizens to the perceived failures of public service delivery, little is known about whether GPAs can also influence the level of government where citizens want policymaking to take place. This question is critical, however, because politicians seeking to impose policy change not only require the backing of the electorate, but also the authority to implement reforms. National-level politicians have more power to implement reforms if authority for policymaking is centralized, whereas subnational-level politicians have more power to implement reforms if authority is decentralized.

In this study, we address whether GPAs themselves shape preferences over which politicians—national or local—have jurisdiction over policymaking. We argue that GPAs should increase preferences for policy centralization. The reason is that benchmarking failure relative to other countries makes citizens more inclined to see policy underperformance as a "national problem requiring national solutions." We claim that citizens should be particularly likely to prefer policy centralization when underperformance is framed as a threat to a country's security. This is because, when people are scared amid crises of confidence, they should "rally 'round the flag" in favor of national unity.

To test our theory, we fielded a survey experiment on a convenience sample of U.S. residents that primed respondents to think about the United States underperforming on education. We randomly varied both the source of information detailing underperformance and how it was presented. Some respondents were given information from a GPA with country comparisons. Others were given information from an NPA without country comparisons. Simultaneously, some respondents were informed why poor test scores should scare them given the importance of education to the United States maintaining geopolitical power. Others were informed why this should cause shame for the country given its wealth.

As predicted, we found that respondents who received the GPA-scare treatment were most likely to prefer policy centralization, whereas respondents who received the NPA-shame control were most likely to prefer policy decentralization. Overall, our findings contribute to a growing literature on how GPAs shape domestic politics. While most studies focus on how GPAs influence demands for policy reform, we demonstrate that GPAs can also affect preferences over what level of government conducts policy. To the extent that GPAs become increasingly prevalent, they could eventually serve to readjust the balance of power between national and subnational politicians over public service delivery.

A natural extension of our study would be to look at the impact of GPAs on other policy areas besides education, such as the environment, healthcare, or transportation. Scholars might also replicate our experiment outside the United States, particularly in contexts where schooling is historically more the responsibility of national, rather than local, politicians. Researchers could additionally address how providing disaggregated, subnational performance data might affect voter preferences for decentralizing or centralizing education. Highlighting which regions, states, or localities do well (and which do not) could condition preferences for where citizens want responsibility for schooling to be vested.

One priority area to explore is whether preferences for policy centralization under GPAs actually prompt strategic responses by politicians. It is plausible, for example, that national-level politicians use the centralizing effects of GPAs to expand their power. Evidence suggests that the "PISA shock" of underperformance has been leveraged in Brazil, Germany, and Japan to push through unpopular and previously untenable education reforms (Rothman, 2017; Takayama, 2008; Volante, 2015). In all three cases, national leaders and bureaucrats used PISA results as coverto wrest power from local-level actors (in the cases of Brazil and Germany; Rothman, 2017) or from predecessors (in the case of Japan; Takayama, 2008).

Ultimately, evidence suggests that GPAs can have significant impacts on the domestic politics of countries. Although international organizations may have originally devised GPAs as a method to improve accountability by discerning which nations are succeeding in policymaking and which are falling behind, their actual effects may be much more wide-ranging. This points to the importance of understanding how GPAs both shape political behavior by the public and the responses by government leaders. Our results demonstrate that GPAs may affect how citizens react to their government in unintended ways, such as influencing citizen preferences over where policymaking authority should be vested.

ACKNOWLEDGMENTS

The authors thank *Governance* editor Adam Sheingate, as well as two anonymous reviewers, for helpful comments and suggestions. They are also grateful to Benjamin Barber IV, Nina Brooks, Leslie K. Finger, and Ken Scheve for providing feedback on the piece. An earlier draft of this article was presented in 2019 at a workshop at IE University, at a workshop at the UCL Institute of Education, and at the Southern Political Science Association's annual meeting. Davies thanks the Hoover Institution at Stanford University for institutional support during the writing of this manuscript. Gift thanks the London School of Economics US Centre and the Stanford Department of Political Science for institutional support. Research for this article was approved by the Human Research Protection Program at Harvard University (protocol IRB17-1901). Data and code necessary to reproduce the numerical results in the article are available in the Harvard Dataverse (https://doi.org/10.7910/DVN/MT38TG).

ORCID

Emmerich Davies https://orcid.org/0000-0002-9686-9960
Thomas Gift https://orcid.org/0000-0001-8684-5677
Carlos X. Lastra-Anadón https://orcid.org/0000-0003-0748-6309

ENDNOTES

[1]GPAs are sometimes called Cross-National Assessments (CNAs) or Global Performance Indicators (GPIs). In education, GPAs are known as International Large-Scale Assessments (ILSAs).

[2]However, Kelley (2017) observes that GPAs can prompt a reshuffling among government officials in terms of who takes responsibility for tasks bearing on policy performance data.

[3]Toch, Thomas. November 30, 2011. "National Problems Require National Solutions." *National Journal*. Available from: http://www.thomastoch.com/wp/2011/national-problems-require-national-solutions/

[4]One of the intents of GPAs is to identify best practices in high-performing countries and to replicate them in countries that perform less well. Yet even when a global consensus can be reached that certain policies are effective, the actual reforms are typically not coordinated internationally, but by national governments.

[5]Ripley, Amanda. December 6, 2016. "What America Can Learn From Smart Schools in Other Countries." *New York Times*. Available from: https://www.nytimes.com/2016/12/06/upshot/what-america-can-learn-about-smart-schools-in- other-countries.html

[6]Vaznis, James. April 10, 2018. "Massachusetts Students Top National Math and Reading Exams." *Boston Globe*.

Available from: https://www.bostonglobe.com/metro/2018/04/09/massachusetts-students-top-national-math-reading- exams/gNR6z4YeNQgBK1MkmFpEBK/story.html

[7]This resembles the intuition in Morgan and Taylor Poppe (2012, p. 264), who activate "fear" by priming respondents to think about the United States's greatest economic rival and then testing how this affects perceptions of local schools. The authors, however, do not explicitly examine the impact of GPAs.

[8]Indeed, wealth may not be the only predictor of whether countries feel embarrassed by poor outcomes on GPAs. In the case of education, for example, even less advanced countries—such as India—have dropped out of PISA, ostensibly to avoid highlighting disappointing results (e.g., Vishnoi, Anubhuti. September 3, 2012, "Poor PISA Score: Govt Blames Disconnect with India." *Indian Express*. Available from: https://indianexpress.com/article/news-archive/web/ poor-pisa-score-govt-blames-disconnect-with-india/). Kelley (2017, pp. 3–8) also provides three vignettes of officials in Israel, Jamaica, and Oman being embarrassed by their relative rankings in a Trafficking in Persons GPA.

[9]Scare framing is especially relevant to policy areas that project strength in the global arena through "hard" or "soft" power. Particularly relevant are performance monitoring devices related to the military, the economy, or technology. Shame framing is germane to nearly any policy area where some countries hold themselves to higher standards. As an example, the release of TIMSS in the United States and PISA in Germany led to framing similar to what we call "scare" and "shame." The media, politicians, and educational researchers used international comparisons to highlight the weaknesses in the education system (Steiner-Khamsi, 2003, p. 2).

[10]Jackson, Abby. December 12, 2016. "The Poorest 10% of Vietnamese Students Performed Better on an International Exam Than the Average American Teen." *Business Insider*. Available from: https://www.businessinsider.com/poor- countries-outperform-america-pisa-exam-2016-12?r=UK

[11]Ibid

[12]Kristof, Nicholas. January 15, 2011. "China's Education System." *New York Times*. Available from: https://www.ny- times.com/2011/01/16/opinion/16kristof.html

[13]Ibid

[14]While the discourse in the United States surrounding ILSAs was initially quiet (Bieber & Martens, 2011; Steiner- Khamsi, 2003), media coverage of PISA has increased in recent years (Green Saraisky, 2015, p. 135). Educational NPAs such as the NAEP, by contrast, have always received substantial attention (Singer & Braun, 2018).

[15]See, for example, the Trends in International Mathematics and Science Study (TIMSS) and the Progress in International Reading Literacy Study (PIRLS).

[16]For example, both No Child Left Behind, advanced by George W. Bush's administration, and Common Core, advanced by Barack Obama's administration, shifted some education policies to the federal government.

[17]According to Bushaw and Calderon (2014, p. 18), 67–77% of the U.S. public (depending on the year) give the schools their children attend A or B grades, compared with only 17–19% giving A or B grades to schools nationwide. Moreover, according to the same survey, 56% of respondents stated that local boards should have the greatest influence "in deciding what is taught in public schools," (Bushaw and Calderon 2014, p. 18) while 60%

opposed the "centralization" of content by the federal government through the Common Core (Bushaw and Calderon 2014, p. 12).

[18]See Strange, Enos, and Hill (2019) for a thorough description of the DLABSS pool.

[19]Our experiment directly compares GPAs to NPAs, which enables us to isolate the effect of the global aspect of the performance assessment from the fact that there exists any type of performance assessment at all. Another option would simply have been to compare a GPA to a control of no information. Yet this would not permit us to distinguish between whether it is the GPA itself that shapes preferences for policy centralization or the fact that respondents received any kind of performance information, regardless of its source.

[20]After completing the survey, participants were debriefed that the information they read was fictitious.

[21]Covariates are: race (white or not), ethnicity (Hispanic or not), age, income from salary, parental status (having any children), education (college graduate or not), and ideology (separate dummies for self-reporting as being on the left and as being moderate). Ideology is based on the question: "Below is a 7-point scale on which the political views that people might hold are arranged from extremely to the left to extremely to the right. Where would you place yourself on this scale?" We code a dummy for those reporting being on the left (1–3), and one for those reporting being moderate (4), with right being the omitted category (5–7). Models with covariates reduce our sample to 408 respondents, which is driven mostly by missingness on age.

[22]When we use models with different baseline categories that allow us to compare the relative effects of the primes, the coefficient on *GP A* × *Scare* is not significantly different from *GP A* × *Shame*. Nor is the coefficient on *GP A* × *Shame* significantly different from *N P A* × *Scare*. Nor is the coefficient on *N P A* × *Scare* significantly different from *N P A* × *Shame*.

[23]Since we rely on a convenience sample, we also reweight our respondents by their demographic characteristics to simulate those of the U.S. population (using poststratification weighting). As shown in Table S2, the *GP A* × *Scare* prime has the largest positive effect, although it is less precisely estimated. We also report analogous logit models in Tables S3 and S4, with similar results.

[24]The exact questions were: [1] "Do you think that government funding for public K-12 schools should increase, decrease, or stay about the same?"; [2] "Students are often given the grades A, B, C, D, and Fail to denote the quality of their work. Suppose the public K-12 schools themselves were graded in the same way. What grade would you give the public K-12 schools in the nation as a whole?"; and [3] "Do you agree or disagree with the following statement? Public K-12 schools need to be reformed." Questions 1 and 2 are taken from the *Education Next* poll (Henderson, Peterson, & West, 2019).

REFERENCES

Bardhan, P. K., & Mookherjee, D. (Eds.). (2006). *Decentralization and local governance in developing countries: A comparative perspective*. Cambridge, MA: MIT Press.

Bieber, T., & Martens, K. (2011). The OECD PISA study as a soft power in education? Lessons from Switzerland and the US: European journal of education, part I. *European Journal of Education*, *46*(1), 101–116.

Bisbee, J. H., Hollyer, J. R., Rosendorff, B. P., & Vreeland, J. R. (2019). The millennium development goals and education: Accountability and substitution in global assessment. *International Organization*, *73*(3), 547–578.

Breakspear, S. (2012). *The policy impact of PISA*. Paris, France: Organisation for Economic Co-operation and Development.

Bushaw, W. J., & Calderon, V. J. (2014). Try it again, uncle Sam: The 46th annual PDK/Gallup poll of the Public's attitudes toward the public schools. *Phi Delta Kappan*, *96*(1), 8–20.

Carney, R. & Enos, R. D. (2017). *Conservatism and fairness in contemporary politics: Unpacking the psychological underpinnings of modern racism*. NYU CESS Experiments Conference. New York.

Channa, A. (2015). *Decentralization and the quality of education*. UNESCO.

Cooley, A., & Snyder, J. (2015). Conclusion - rating the ratings craze: From consumer choice to public policy outcomes. In A. Cooley & J. Snyder (Eds.), *Ranking the world: Grading states as a tool of global governance*. Cambridge, MA: Cambridge University Press.

Davis, E., Wilson, R., & Dalton, B. (2018). Another slice of PISA: An interrogation of educational cross-national attraction in Australia, Finland, Japan and South Korea. *Compare: A Journal of Comparative and International Education*, *50*(3), 309–331.

Davis, K. (2012). *Governance by indicators: Global power through classification and rankings*. Oxford, MA: Oxford University Press.

Doshi, R., Kelley, J. G., & Simmons, B. A. (2019). The power of ranking: The ease of doing business indicator and global regulatory behavior. *International Organization*, *73*(3), 611–643.

Falleti, T. G. (2010). *Decentralization and subnational politics in Latin America*. Cambridge, MA: Cambridge University Press.

Fiske, E. B. (1996). *Decentralization of education: Politics and consensus*. World Bank: Washington, DC.

Fusarelli, L. D., & Fusarelli, B. C. (2008). Federal Education Policy from Reagan to Obama: Convergence, divergence, and 'control'. In B. S. Cooper & J. G. Cibulka (Eds.), *Handbook of education politics and policy*. New York: Routledge.

Gershberg, A. I., González, P. A., & Meade, B. (2012). Understanding and improving accountability in education: A conceptual framework and guideposts from three decentralization reform experiences in Latin America. *World Development*, *40*(5), 1024–1041.

Goldin, C., & Katz, L. F. (2010). *The race between education and technology*. Cambridge, MA: Harvard University Press.

Gourevitch, P. (1978). The second image reversed: The international sources of domestic politics. *International Organization*, *32*(4), 881–912.

Green Saraisky, N. (2015). *The politics of international large-scale assessment: The Programme for international student assessment (PISA) and American education discourse, 2000–2012*. (PhD Dissertation), Columbia University, New York, NY.

Grek, S. (2009). Governing by numbers: The PISA 'effect' in Europe. *Journal of Education Policy*, *24*(1), 23–37.

Guerra, S. C., & Lastra-Anadón, C. X. (2019). The quality-access tradeoff in decentralizing public services: Evidence from education in the OECD and Spain. *Journal of Comparative Economics*, *47*(2), 295–316.

Healy, A., & Malhotra, N. (2013). Retrospective voting reconsidered. *Annual Review of Political Science*, *16*(1), 285–306.

Henderson, M. B., Houston, D. M., Peterson, P. E., & West, M. R. (2020). Public support grows for higher teacher pay and expanded Schol choice. *Education Next*, *20*(1), 10–27.

Henderson, M. B., Peterson, P. E., & West, M. R. (2019). *Trends in the EdNext Survey: Question wording and data over time*. Available from https://educationnext.org/files/ednext-poll-question-wording- over-time-through-2018.pdf

Heyneman, S. P., & Lee, B. (2014). The impact of international studies of academic achievement on policy and research. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment*, (p. 36). London.

Kelley, J. G. (2017). *Scorecard diplomacy: Grading states to influence their reputation and behavior*. Cambridge, MA: Cambridge University Press.

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.

Kelley, J. G., & Simmons, B. A. (2019). Introduction: The power of global performance indicators. *International Organization*, *73*(3), 491–510.

Kijima, R., & Lipscy, P. Y. (2017). *The politics of international testing*. Stanford, CA: Stanford University.

Lassiter, M. D. (2007). *The silent majority: Suburban politics in the sunbelt south*. Princeton, NJ: Princeton University Press.

Lockheed, M., Prokic-Bruer, T., & Shadrova, A. (2015). *The experience of middle-income countries participating in PISA 2000-2015*. Washington, DC: PISA, World Bank, OECD.

Manor, J., & Crook, R. C. (1998). *Democracy and decentralisation in South Asia and West Africa: Participation, accountability, and performance*. Cambridge, MA: Cambridge University Press.

Merry, S. E., Davis, K., & Kingsbury, B. (2015). *The quiet power of indicators: Measuring governance, corruption, and the rule of law*. Cambridge, MA: Cambridge University Press.

Morgan, S. L., & Taylor Poppe, E. S. (2012). The consequences of international comparisons for public support of K-12 education: Evidence from a national survey experiment. *Educational Researcher*, *41*(7), 262–268.

Mueller, J. E. (1970). Presidential popularity from Truman to Johnson. *American Political Science Review*, *64*(1), 18–34.

National Center for Education Statistics. (2018). *Digest of Education Statistics*, 2018.

Norpoth, H. (1984). Economics, politics, and the cycle of presidential popularity. *Political Behavior*, *6*(3), 253–273.

Papanicolas, I., & Jha, A. K. (2017). Challenges in international comparison of health care systems. *Journal of the American Medical Association*, *318*(6), 515–516.

Pizmony-Levy, O. (2018). Compare globally, interpret locally: International assessments and news media in Israel. *Globalisation, Societies and Education*, *16*(5), 577–595.

Pizmony-Levy, O., & Bjorklund, P. (2018). International assessments of student achievement and public confidence in education: Evidence from a cross-national study. *Oxford Review of Education*, *44*(2), 239–257.

Pizmony-Levy, O., Doan, L., Carmona, J., & Kessler, E. (2017). *The public and international assessments*. Teachers College at Columbia University.

Pizmony-Levy, O., & Torney-Purta, J. (2018). How journalists and researchers communicate results of international large-scale assessments. *Giornale italiano di pedagogia sperimentale*, *26*(1), 51–65.

Putnam, R. D. (1988). Diplomacy and domestic politics: The logic of two-level games. *International Organization*, *42*(3), 427–460.

Ramirez, F. O., Schofer, E., & Meyer, J. W. (2018). International tests, national assessments, and educational development (1970–2012). *Comparative Education Review*, *62*(3), 344–364.

Rodden, J. (2004). Comparative federalism and decentralization: On meaning and measurement. *Comparative Politics*, *36*(4), 481.

Rothman, B. (2017). *After the shock: The German education system in 2017*. National Center on Education and the Economy.

Schleicher, A. (2019). *PISA 2018: Insights and interpretations*. Paris, France: Organisation for Economic Co- operation and Development.

Sides, J., Tesler, M., & Vavreck, L. (2019). Identity crisis: The 2016 presidential campaign and the battle for the meaning of America (2nd ed.). Princeton, NJ: Princeton University Press.

Singer, J. D., & Braun, H. I. (2018). Testing international education assessments. *Science*, *360*(6384), 38–40.

Stack, M. (2007). Representing school success and failure: Media coverage of international tests. *Policy Futures in Education*, *5*(1), 100–110.

Steiner-Khamsi, G. (2003). The politics of league tables. *JSSE-Journal of Social Science Education*, *1*, 6.

Strange, A. M., Enos, R. D., & Hill, M. (2019). Online volunteer laboratories for human subjects research. *PLoS One*, *14*(8), e0221676.

Takayama, K. (2008). The politics of international league tables: PISA in Japan's achievement crisis debate. *Comparative Education*, *44*(4), 387–407.

Tampio, N. (2017). Democracy and national education standards. *The Journal of Politics*, *79*(1), 33–44.

Tiebout, C. M. (1956). A pure theory of local expenditures. *The Journal of Political Economy*, *64*(5), 416–424.

Tversky, A., & Kahnemann, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Volante, L. (2015). The impact of PISA on education governance: Some insights from highly reactive policy contexts. *International Studies in Educational Administration*, *43*(2), 103–117.

Wibbels, E. (2006). Madison in Baghdad?: Decentralization and federalism in comparative politics. *Annual Review of Political Science*, *9*(1), 165–188.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Davies E, Gift T, Lastra-Anadón CX. How global performance assessments shape attitudes toward government decision-making: Survey experimental evidence. *Governance*. 2020;1–21. https://doi.org/10.1111/gove.12504