The Politics of International Testing

Rie Kijima[1]
Graduate School of Education
Stanford University

Phillip Y. Lipscy
Department of Political Science and the Freeman Spogli Institute for International Studies
Stanford University

In recent years, an increasing number of countries have participated in cross-national assessments in education (CNAs), but their impact remains underexplored. We argue that CNA participation increases the capacity and motivation of policymakers to implement improvements in education through mechanisms at the elite, domestic, and transnational levels. We find evidence consistent with our propositions using a mixed-method approach, utilizing: 1. a panel dataset covering all CNAs and all countries in the international system; 2. an original survey of 77 education officials directly responsible for the planning and implementation of CNAs in 46 countries; 3. personal interviews with 48 officials in target states, assessment agencies, and donor agencies. The empirical results consistently support our prediction that CNA participation has a meaningful impact on education policy and outcomes.

[1] Rie Kijima (Stanford University) is Interim Director of the Master's Program in International Comparative Education/International Education Policy Analysis, Graduate School of Education, and Phillip Y. Lipscy (Stanford University) is Assistant Professor of Political Science and Thomas Rohlen Center Fellow at the Freeman Spogli Institute for International Studies. Paper prepared for the "Assessment Power in World Politics" conference hosted by Harvard University and *International Organization*, May 6-7, 2016 and the associated APSA mini-conference, September 2, 2016. We would like to thank Felipe Barrera-Osorio, Eric Bettinger, Patricia Bromley, Martin Carnoy, Judith Goldstein, Elena Tej Grewal, Edward Haertel, Judith Kelley, David Laitin, Aila Matanock, John Meyer, Lambrina Mileva-Kless, Sonal Pandya, Francisco Ramirez, Kenneth Schultz, Beth Simmons, Michael Tomz, Catherine Weaver, Imeh Williams, and conference participants at APSA 2016, Harvard, Stanford, and U.C. Berkeley for their valuable feedback. Jane Leer, Idalia Rodriguez Morales, and Trevor Incerti provided excellent research assistance. We thank all individuals who responded favorably to requests for interviews, survey completion, and documents collected over a period of three years. We also thank Dr. Hans Wagemaker, former Director of International Association for the Evaluation of Education (IEA), and Dr. Dirk Hastedt, current Director of IEA for their unwavering support of this study. This research was financed by the Stanford Global Development and Poverty Initiative, Stanford Graduate School of Education, Education International's Mary Futrell Scholarship Fund, the Freeman Spogli Institute for International Studies at Stanford University, the Center of East Asian Studies at Stanford University, and Stanford Center on Philanthropy and Civil Society.

The number of countries participating in cross-national assessments in education (CNAs) has grown rapidly over the last 50 years. Countries are increasingly willing to use assessments to measure and disseminate the state of their education. This represents a stark departure from the traditional politics of education, in which information about student performance was largely contained within the borders of nation-states.[2] Assessment results from CNAs such as Trends in International Mathematics and Science Study (TIMSS) and the Program for International Student Assessment (PISA) are now widely followed as indicators for national educational quality and, more broadly, human capital and international competitiveness.[3] More than 60 countries now regularly participate in CNAs, a fivefold increase from 1959. Participation among developed countries is now nearly universal, and about half of participants are developing countries.[4]

CNAs are an important substantive topic of inquiry for several reasons. First, by allowing countries to benchmark their progress over time and against their peers, CNAs have the potential to improve education performance and economic outcomes. Education policy directly affects the prospects of over 1.8 billion school-age children in the world today, with more to come.[5] In addition, education quality has been widely recognized as an important source of economic development.[6]

Second, CNAs have been the subject of considerable controversy and value contestation. At their best, CNAs bring transparency to policymaking by offering a standard metric for cross-national comparison. They can help international and domestic audiences hold political leaders

[2] Anderson 2006.
[3] Among others, see "Testing Education: Pisa Envy," *The Economist*, 1-19-2013.
[4] Based on data from TIMSS. We use the World Bank classification of economies to define developing countries, which are low-income, lower-middle income, and upper middle income countries.
[5] World Bank Group 2011.
[6] Hanushek and Kimko 2000; Hanushek and Woessmann 2012; Hanushek and Woessmann 2008; Rodrik 1995.

accountable for education quality and performance.[7] However, critics have argued that test scores are often misinterpreted, leading to misleading conclusions.[8] Others have condemned CNAs as "educational colonialism,"[9] which shifts the focus of education towards "teaching to the test" [10] and away from less quantifiable goals such as personal and moral development.[11]

Finally, the effect of CNAs on policy remains understudied by social scientists. Existing research has predominantly focused on the role of global norms and culture institutionalized by Western countries and multilateral agencies,[12] which have compelled transitional and developing countries to participate in CNAs.[13] Although this is a useful framework for explaining the general proliferation of CNAs, it is less useful for explaining their impact. Studies that consider the influences of CNAs on policymaking have largely focused on case studies[14] – to the best of our knowledge, this is the first systematic study of the impact of CNAs on education policy and outcomes.

In this article, we will consider the "assessment power"[15] of CNAs. Do CNAs affect education policy and education outcomes, and if so, how? We argue that CNAs increase both the capacity and motivation of policymakers to implement improvements in education. CNAs exert a particularly strong influence at the elite level. Because administering a CNA requires extensive elite interaction with authoritative agencies and experts, there is considerable scope for technical transfers, learning, and socialization. CNAs also generate detailed information about the

[7] Benveniste 1999; Carnoy 2014.
[8] Loveless 2012; Koretz 2009.
[9] "OECD and Pisa tests are damaging education worldwide – academics," *The Guardian*, 5-6-2014, https://www.theguardian.com/education/2014/may/06/oecd-pisa-tests-damaging-education-academics
[10] Jennings and Bearak 2014; Sutton 2004; Volante 2004.
[11] Meyer and Benavot 2013.
[12] Ramirez, Meyer, and Lerch 2015; Smith 2016.
[13] Kamens and McNeely 2010.
[14] Abdul-Hamid, Abu-Lebdeh, and Patrinos 2011; Takayama 2008; Addey 2015; Grek 2009.
[15] Kelley and Simmons, this issue

shortcomings of a country's education system, which can inform reforms. In addition, the clear, transparent, and comparative rankings produced by CNAs invoke strong status concerns among elite policymakers that serve as a motivation for improvements in education policy. We expect domestic and transnational pressures for reform to be conditional: public pressure generated by poor rankings is an important motivation for education reforms among more democratic states, and international pressure, particularly from donor agencies, is an important impetus for aid-dependent developing countries.

Empirically, we adopt a mixed-methods approach. First, we analyze a panel dataset covering all CNAs and all countries since 1950. The panel analysis provides strong evidence that CNA participation is associated with increases in net secondary enrollment and education aid inflows. We use several empirical strategies to address potential endogeneity, i.e. the possibility that preexisting education reforms are responsible for both CNA participation and changes to education outcomes. Second, many of the pathways of influence associated with our proposed causal mechanisms are either difficult to observe or quantify (e.g. the relative status perceptions of policymakers; learning and knowledge transfers). To address this challenge, we conducted an elite survey of 77 education officials directly involved in the planning and implementation of CNAs in their countries. The survey allows us to examine the perceived importance of specific pathways of influence that would otherwise be impossible to observe. In addition, we personally interviewed 48 policymakers in both target states and assessment agencies to qualitatively examine the assessment power of CNAs. The quantitative and qualitative evidence provide consistent support for our proposition that CNA participation has a substantial impact on education policy and outcomes.

The Rise of Cross-National Assessments in Education

The first CNA, the Pilot Twelve-Country Study, was administered in 1959. Western countries were riveted by the launch of Sputnik and attributed the success of the Soviet space program to high-quality science education.[16] Educational researchers expressed concerns about their inability to judge the quality of education cross-nationally. Driven by a strong desire to demonstrate educational progress empirically, a group of mostly Western industrialized countries joined the effort to create an "internationally valid standard."[17] The Twelve-Country Study consisted of 120 test items that covered reading, math, science, and geography.[18]

CNAs encompass both global assessments and regional assessments. Global assessments, like TIMSS and PISA, are universalistic in spirit. Although participation is voluntary, and hence universality has not been achieved, global assessments place few restrictions on participation based on country-specific factors.[19] Regional assessments are administered in countries from a particular region, such as Africa or Latin America. Examples of regional assessments are the Southern African Consortium of Education Quality (SACMEQ) in Africa and the Second Regional Comparative and Explanatory Study (SERCE) in Latin America.

Today, over 60 countries and economies regularly participate in global international assessments like PISA, TIMSS, and the Program in International Reading and Literacy Study (PIRLS) conducted by the International Association for the Evaluation of Educational

[16] Husén 1979, 374.
[17] Ibid.
[18] Husén 1979
[19] Countries need to implement the assessments themselves in cooperation with assessment agencies, and CNAs generally involve a membership/participation fee. However, financial and technical assistance is usually available from donor or assessment agencies even for less economically developed countries. As a reference, the cost associated with TIMSS Grade 8 assessment is approximately 40,000 USD per year, excluding costs associated with analysis and dissemination Greaney and Kellaghan 2008, vol. 1, 75. OECD has a fee structure that corresponds with the country's economic development (author interview with an education expert at OECD, 2012).

Achievement (IEA). Figure 1 traces participation in CNAs since 1959. CNAs primarily attracted a handful of economically advanced nations from 1959 through the 1980s. Participation expanded dramatically since the 1990s, particularly among less developed countries.[20] In 2015, 72 countries and economies participated in PISA.[21] With each round, international assessments in education have attracted more participants from around the world.

The Impact of Cross-national Assessments in Education

While the proliferation and increasing visibility of CNAs is unmistakable, very little scholarship has systematically examined how CNAs affect policy outcomes. Our central argument is that CNA participation increases the capacity and motivation of education policymakers to pursue improvements in education quality. CNA participation improves the capacity of education officials to implement effective reforms by increasing access to technical expertise, training, and information about shortcomings in their education systems. In addition, CNA participation increases the motivation of policymakers to pursue improvements in education quality by generating status competition, socialization, and domestic and international pressure for reform. In this section, we will discuss the specific features of CNAs and how they affect education policymaking and outcomes.

It is helpful to place CNAs within the broader context of global performance assessments (GPAs), "regularized, publicized reporting routines that states, IGOs, NGOs, or private actors use to attract attention to the relative performance of countries or other organizations in a given policy

[20] There is a jump in the number of countries participating in CNAs in 1990, which falls on a year when three CNAs were administered at the same time. These three tests are: International Assessment of Educational Progress, Reading Literacy Study, and Pacific Islands Literacy Level.
[21] OECD, "About PISA," https://www.oecd.org/pisa/aboutpisa/

or performance area."[22] CNAs clearly fit the criteria for a GPA. The results of CNAs are generally *public* and presented openly, clearly, and simply, resulting in transparency and wide dissemination.[23] Major CNAs are conducted on a *regular* and predictable cycle.[24] CNAs are *inclusive* and explicitly *comparative* in nature, seeking to rank countries based on student performance in subject areas such as reading, math, and science. Assessment agencies conduct CNAs in a *purposive* manner, linking test results to improvements in education policy.[25] Participation in a CNA thus publicly reveals information about a country's education performance and its relative standing in international comparison.

CNAs generally involve extensive interaction and feedback between education policy elites in the target country and authoritative assessment agencies. This makes elite mechanisms a particularly important source of assessment power. Unlike GPAs that rely primarily on existing data sources,[26] CNAs generate new data through the active participation of students, teachers, school administrators, and government officials in the target country. Hence, CNAs are typically conducted by countries in close cooperation with assessment agencies, which provide direct and extensive assistance with development of test items, planning, sampling, field trials, and analysis.[27] Other organizations, such as bilateral and multilateral donor agencies, often assist with funding

[22] Kelley and Simmons, this issue

[23] There is some variation in transparency regarding specific data associated with CNAs. For example, while the headline national test scores are disseminated widely, and data by gender, socio-economic status, and performance variance are readily available, some potentially useful data to conduct longitudinal analysis are not released to the public (e.g. https://www.oecd.org/pisa/aboutpisa/pisafaq.htm).

[24] E.g., PISA is conducted every three years, and TIMSS every four years.

[25] E.g., the IEA website's overview of TIMSS notes that "Assessing fourth grade students can provide an early warning for necessary curricular reforms, and the effectiveness of these reforms can be further monitored at the eighth grade." (http://www.iea.nl/timss); OECD similarly highlights how PISA shapes education reform, noting that "countries such as Germany and Brazil have been able to improve their student performance in PISA and make their education systems more inclusive." (http://www.oecd.org/pisa/aboutpisa/)

[26] e.g. the Aid Transparency Index is produced without significant consultation with target governments as discussed by Honig 2016; Weaver 2016.

[27] Henry et al. 2001; Lockheed 2012.

and administration of the assessment and policy design in response to the findings.[28] It is also common for governments to involve domestic and international academics and education policy experts to provide technical support at all stages. Education policymakers from participant countries also attend conferences and receive extensive information about successful reforms implemented in other countries.[29] Hence, CNAs provide significant opportunities for capacity building of education policy elites through interaction with assessment agencies, international organizations, foreign counterparts, and technical experts.[30]

CNA participation is not just an exercise in capacity building. Government interaction and learning surrounding CNAs involves authoritative counterparties with strong reputations and technical capabilities.[31] OECD, the administrator of PISA, draws on its prestige as a club of the most advanced economies in the world as well as its research and technical capabilities.[32] Similarly, IEA draws on its expertise as a pioneer in international assessments and nearly 60-year history of conducting CNAs worldwide.[33] Since the end of WWII, international aid agencies have played a central role in setting the agenda on global governance in education.[34] Today, these organizations play a critical role in constructing "shared ideologies of an 'imagined' world order through process of negotiation, diffusion, and sometimes contestation."[35] Assessment and aid agencies leverage their authority and access to education policy elites to motivate changes in

[28] Kijima 2010.
[29] For example, a 2013 IEA Conference for education policymakers included presentations on technical issues such as "Methodological Issues in Large-Scale Assessment" and "Using HLM With International Large-Scale Assessment Data," as well as presentations addressing the content of education, such as "Mathematics Education in Singapore," "Predictors of Science Achievement," and "Quality of Schools and Teaching: What Can We Learn From International Studies?" (5th IEA International Research Conference 26–28 June 2013, Conference Program).
[30] Lockheed 2012.
[31] This is similar to arguments made in this issue regarding the World Bank and International Labour Organization. See Kelley, Simmons, and Doshi 2016; Koliev, Sommerer, and Tallberg 2016.
[32] Meyer and Benavot 2013.
[33] Mullis and Martin 2007, 9.
[34] Mundy 1999; Mundy and Manion 2014.
[35] Mundy 1999, 28.

policies consistent with global values such as "education for all," gender equality, and an emphasis on economically important skills such as reading, math, and science.[36]

CNA participation also motivates education policy elites by generating status competition. CNAs reveal, for all to see, quantitative information about a country's education performance and relative standing in international comparison. Education is a fundamental policy issue that touches upon essentially all citizens, and it is also naturally associated with status competition as the first institutionalized setting that subjects children to evaluation and relative comparison. For some countries, education performance is deeply tied to a sense of identity and self-worth, while for others, it is perceived as a proxy for broader economic or geopolitical competition.[37] Assessment agencies also reinforce status competition by sponsoring conferences that routinize personal, cross-national interaction among education policy elites of participant countries.[38] We thus expect CNA participation to facilitate status competition and strengthen the motivation of policy elites to make improvements to their education systems.

Compared to other GPAs, we expect transnational pressure to be somewhat less salient and conditional for CNAs. In contrast to GPAs such as terrorist designations by the US government,[39] assessment agencies do not explicitly link CNAs to formal rewards or punishments: there is no direct channel for international pressure. However, CNAs provide reliable data on the status and progress of a country's education system. Hence, international donor agencies may be more willing to disburse aid to CNA-participating countries, where the scores can serve as benchmarks for the monitoring and evaluation of projects. In addition, CNA scores may also serve as a proxy for quality of human capital, influencing decisions by private investors. Insofar as these

[36] Martens and Niemann 2013.
[37] Ibid.
[38] See Chabbott 2003 for an analysis of Education for All Conferences.
[39] Morse 2016; Jo, Phillips, and Alley 2016.

informal transnational pressures exist, we expect them to be particularly salient for developing countries that are dependent on aid and investment flows.

We also expect the observed impact of domestic political pressures to be somewhat weaker and conditional for CNAs compared to other GPAs. Low CNA performance can trigger public criticism and upheaval, facilitating education reforms.[40] However, several factors may limit the impact of domestic political mechanisms. First, domestic political pressure is likely to have a much greater impact in democratic states, where leaders face electoral repercussions from poor performance. In autocratic states, which have been participating in CNAs in greater numbers, leaders may be more able to ignore or suppress domestic pressures. Second, education reforms are often highly politicized – e.g. charter schools, higher salaries for teachers – which may limit the ability for domestic groups to rally around a coherent objective.[41] More broadly, unlike GPAs that target government policies or activities,[42] which can be reformed directly by executive or legislative action, governments exercise only indirect control over the academic performance of schoolchildren.[43] This is not to say that education outcomes are immovable: evidence from impact evaluations indicate that there many practical policy interventions available to policymakers that rapidly and cheaply move salient education outcomes, such as student and teacher absenteeism.[44]

Third, participation in CNAs is largely voluntary. The voluntary nature of participation raises the potential problem of self-selection. Specifically, leaders who particularly fear the domestic political repercussions of publicizing the state of their education systems may opt out,

[40] Takayama 2010.
[41] e.g. Ladd 2002.
[42] Some GPAs explicitly target actions directly under government control, e.g. the Aid Transparency Index (Honig 2016; Weaver 2016). and subnational performance indicators focusing on administrative procedures (Le and Malesky 2016).
[43] Analogously, governments may find it difficult to reduce corruption even when they seek to do so (Lee and Matanock 2016).
[44] Kremer, Brannen, and Glennerster 2013. Miguel and Kremer 2004. Benhassine et al. 2013. Duflo, Hanna, and Ryan 2012.

limiting the capacity of CNAs to shape the policies of the most problematic countries.[45] As we will illustrate below, this is not always a problem: leaders sometimes opt for participation despite low expected performance, and some countries participate in CNAs due to external mandates. Self-selection also introduces nontrivial methodological challenges in evaluating the impact of CNAs, which we discuss in the empirical section.

In summary, we predict that CNA participation will have tangible effects on education policymaking and facilitate improvements in education quality. Concerning mechanisms, we expect *elite politics* to be particularly important for CNAs. Relative to other GPAs, CNAs inherently involve extensive interaction between target states on the one hand and assessment agencies, assessment experts, and aid agencies on the other hand. This suggests that CNAs are highly likely to provide significant opportunities for policy change through elite mechanisms such as learning, socialization, and status concerns. We expect the role of *transnational politics* to be conditional, as CNAs are not associated with explicit rewards or sanctions. However, informal rewards, such as increases in foreign aid or investment, are likely to be an important consideration for policymakers in less developed countries. The importance of *domestic politics* as a mechanism is also likely to be conditional. The simplicity, transparency, and comparability of test scores provide ample opportunity for the mobilization of public opinion and civil society groups around education reform. However, the domestic political channel is likely to be limited in less democratic societies. In addition, the observable efficacy of the domestic political channel may be mitigated due to politicization of education reforms and self-selection: political leaders particularly concerned about domestic repercussions may opt out ex ante.

[45] Kijima 2013.

Panel Analysis

In this section, we will use a panel dataset to examine the impact of CNA participation on education quality. To do so, we constructed an original dataset covering all countries in the international system from 1950 to 2012. The dataset contains participation status by country for all CNAs, including major cross-national assessments and smaller-scale regional assessments. A full list of CNAs included in the panel is available in Table 1.

CNAs provide the best available measure for national education quality. One observable implication of our theory is that, *ceteris paribus*, CNA scores should be higher for countries that initiated participation in CNAs earlier, assuming enhancements to education quality accumulate over time. This is indeed the case: the correlation between average test scores in 2005 and number of years since first CNA participation is 0.43.[46] However, a major limitation of the CNA data is that it is unavailable for non-participating states and participating states prior to their first test. Hence, it is impossible to rule out obvious alternative explanations such as self-selection: countries with better education quality likely selected into assessments early on. This necessitates an alternative measure with greater cross-national and temporal availability.

We use net secondary school enrollment as a proxy for education quality and attainment.[47] Net secondary enrollment is a reasonable alternative measure for several reasons. First, the measure is widely used in existing studies as a proxy for educational development and attainment.[48]

[46] This association is robust to controlling for GDP/capita and regime type. For CNA scores, we use data from Angrist, Patrinos, Schlotter (2012) that unified measure of cognitive achievement by connecting scores from various CNAs between 1965-2010.

[47] Source: World Development Indicators, World Bank, unit is %.

[48] Scholars have used secondary enrollment rate to proxy for both access and quality of education. For example, scholars employ secondary enrollment rate to proxy for educational development which leads to higher human capital, economic growth, and greater innovation Vandenbussche, Aghion, and Meghir 2006; Dorius 2013; Delaney and Yu 2013.

This is intuitive: increases in secondary enrollment tend to reflect stronger student performance at the primary level or policy initiatives to enhance access to higher levels of education. Secondary education attainment is widely cited as essential for the transition from school to the labor market in a knowledge-based economy, and many countries include secondary enrollment in benchmarking the progress of their education system.[49] Second, unlike primary enrollment, which is universal in a large majority of countries, there is meaningful cross-national and temporal variation in net secondary enrollment rates.[50] Third, most CNAs in our dataset, including those with the largest number of country participants, such as TIMSS and PISA, target youths in upper levels of primary and lower secondary education.[51] As such, reforms motivated by CNAs will most likely have an impact at this level. Finally, among various potential education measures that we considered, net secondary enrollment exhibits the strongest correlation with available CNA scores (R=0.67).[52]

The key independent variable is *CNA participation*. This is a dichotomous variable coded as 0 for countries that have never participated in a CNA, and 1 after a country initiates participation for the first time. We use this coding scheme for several reasons. First, we view the first instance of participation in a CNA is a critical milestone. The first instance of participation generates a country-specific score and relative ranking for the first time. In addition, switching from non-

[49] Bloom 2006; UNESCO Institute for Statistics 2010; Stone et al. 2013.

[50] E.g., even for developed countries that initiated participation relatively late in our data, net secondary enrollment was not maxed out and had room to grow, e.g. Denmark in 1990 was at 87%.

[51] TIMSS is administered to 4th (upper elementary) and 8th (lower secondary) school-aged students; PISA is conducted to 15 year olds, equivalent to first or second year of upper secondary education

[52] For CNA scores, we use data from Angrist, Patrinos, and Sclotter 2013, who constructed a unified measure of cognitive achievement by connecting scores from various CNAs between 1965-2010. Another plausible measure for education quality is government spending on education. However, the correlation between education spending as a share of GDP and CNA scores is very low (R = 0.08). There were analogous problems with percentage of secondary teachers who are trained (R=0.05) and duration of compulsory education (R=0.17). Other variables for educational attainment were considered but rejected due to limited availability, such as the attainment indicator by Barro and Lee (2013) which is available in 5-year intervals, and Cohen and Soto's (2007) educational attainment data compiled every 10 years.

participation to participation triggers several developments that could plausibly have a large impact on education policymaking – intensive elite interaction with assessment agencies for the first time, as well as the first revelation of a country's score and relative ranking to policymaking elites and domestic and international audiences. While subsequent participation will also plausibly produce an incremental impact, the clearest test of our theory is the first shift from non-participation to participation. Second, CNAs are not usually conducted annually, and coding off years as non-participation is likely to produce misleading results. Third, it is impractical to code "dropouts" from CNAs because of the irregular timing of CNAs and the fact that we do not have a long track record since initiation of participation for most countries.[53] Our theoretical priors also suggest dropping out is unlikely to have a "reverse effect" of the same magnitude as participating for the first time – once a country participates in a CNA, a score and relative ranking will be available regardless of future behavior, and we would not expect the learning and norm diffusion that took place during the initial round of participation to be reversed by a decision to drop out. We thus divide countries broadly into two categories, i.e. "non-treated" countries that have never participated in a CNA and "treated" countries that have.

For the purposes of empirical analysis, it is important to be cognizant of the logistics of CNA administration. Figure 2 depicts a typical timeline for a country that participates in a CNA.[54] A country will typically declare their intention to participate in a CNA about 2-3 years in advance of the assessment date. The period immediately preceding the assessment involves preparations

[53] It is easy to observe non-participation for a given round of a particular CNA, but countries often temporarily cease participation in one CNA while prioritizing another CNA.

[54] The specific values in the figure are based on PISA 2015. We were unable to collect comprehensive data regarding the timing of events for each CNA in our dataset due to unavailable records for older CNAs. However, we obtained the relevant dates for 23 CNAs, and the only major exception to the timeline we found was the 2008 Teacher Education and Development Study in Mathematics administered by IEA, which published results with an unusually large lag (four years).

and interaction with the assessment agency and other relevant parties, such as donor agencies. Test scores are typically released about a year after the assessment takes place. This means that the effect of participation on education policymaking will tend to occur in two distinct phases, which gives us some leverage over causal mechanisms: 1. Primarily elite interaction with international actors in the run up to the assessment, which provides opportunities for learning, professionalization, and socialization; 2. Domestic and international response to the release of test scores after administration of the assessment.

For the purposes of empirical analysis, we will incorporate some flexibility regarding the specific timing of CNA initiation to account for these factors. It is also important to recognize that policy measures in response to participation are unlikely to take effect immediately, i.e. we would not expect secondary enrollment to increase instantaneously upon the initiation of consultations with assessment agencies or the revelation of test scores.

For our analysis, we use a generalized difference-in-differences estimation, controlling for year and country fixed effects across all OLS specifications. We also control for GDP/capita (PPP) to account for differences in education outcomes associated with levels of development and polity scores to account for the fact that democracies tend to invest greater resources in public goods, such as education.[55]

In Table 2, we examine the association between CNA participation and secondary enrollment rates. CNA participation is positively associated with secondary school enrollment across specifications. In the second column, we include foreign aid to education as a control variable to examine whether the effect of participation on secondary enrollment runs largely through shifts in education aid. The results indicate that education aid is not strongly associated

[55] Baum and Lake 2003.

with secondary school enrollment. Even after controlling for education aid, the association between participation and enrollment remains positive and statistically significant, suggesting that other mechanisms are likely at work. As we discussed earlier, the logistics of CNAs are such that effects on policy outcomes could somewhat lag behind administration – e.g., this would be the case if the main impact of CNAs occurs from policy responses to the publication of test scores. Alternatively, we might conceptualize participation differently and treat expressions of intent to participate as the true initiation of participation. To account for these possibilities, we reran our empirical specifications by leading and lagging the key independent variable by three years.[56] As the third and fourth column of Table 2 illustrate, this produces substantively similar results.

An important assumption of our empirical approach is that CNA participants and non-participants would have been subject to common trends in secondary enrollment in the absence of CNA participation. If CNA participants exhibit a higher rate of increase in secondary enrollment prior to participation, it would be strongly indicative of endogeneity: e.g. countries with preexisting education reform efforts selecting into CNAs. Following the approach of Artur,[57] we reran our empirical specifications including indicator variables for leads and lags of participation to examine if the participant group exhibits any distinct trends in secondary enrollment prior to participation. More specifically, we omit our key independent variable and instead include dummy variables for t-4, t-3, t-2, t-1, t=0, t+1, t+2, t+3, and t>3, where t=0 is the first time a country participated in a CNA. Each indicator variable is coded as one only in the relevant year, with the exception of t>3, which is coded 1 for all years subsequent to year 3. The substantive results are presented in Figure 3. The figure illustrates that, among countries that participated in CNAs, there

[56] One or two years produce substantively similar results.
[57] Artur 2003.

is no trend in secondary enrollment prior to the year they initiated participation, i.e. there does not appear to be any anticipatory increase in education quality prior to CNA initiation.

In order to interpret the findings in Figure 3, it is helpful to refer back to the timeline we presented in Figure 2. Elite interaction with international actors generally commences about two years prior to the administration of an assessment. Strictly speaking, this means that it would not be contrary to our theoretical propositions if we observed some shift in secondary enrollment between t-2 and t=0 (it would be far more problematic if we observed a trend to prior to t-2). However, it is unrealistic to expect that policy measures to boost secondary school enrollment would be developed and implemented immediately. The observation of an effect on secondary enrollment at t=0 is broadly consistent with the proposition that CNAs exert an effect on education outcomes through elite politics. Another possibility is that knowledge of upcoming CNA score revelation motivates education officials to implement policies to improve learning outcomes, such as measures to reduce student and teacher absenteeism. Policy reforms in response to the actual revelation of test scores will generally occur after t+1, when scores are typically released. Figure 3 shows that secondary enrollment continues to climb after t+1, though it is not possible to distinguish whether this is due to scores becoming public or the continuation of earlier effects. At a minimum, our results suggest that the impact of CNAs on enrollment is not due *entirely* to events that unfold after the publication of test scores. This is consistent with our theoretical priors that elite politics are a particularly important source of influence for CNAs. We will discuss causal mechanisms in greater detail in the next section.

Although we have shown that there is no difference in the pre-trend for CNA participants and non-participants, one residual concern is self-selection in the period immediately surrounding CNA participation. More specifically, some countries might initiate participation in CNAs as part

of a new education reform policy: CNAs may be seen as a way to better evaluate the effect of reforms. Under these circumstances, we could potentially observe CNA participation coinciding with improvements in education quality even if CNA participation per se has no effect.

To address this concern, we classify countries according to whether or not CNA participation was triggered by domestic education reform. We are able to do this because one of the questions we administered in our survey explicitly asked officials if domestic reforms were an important reason for their country's first-time participation in CNAs.[58] We distinguish countries that answered in the affirmative and negative to this question.[59]

Based on our identification of reform and non-reform countries, we recoded our key independent variable, CNA participation. As with the original variable, countries are coded as 1 for all years after participation is initiated, subject to being a *non-reform country*. Using an analogous procedure, we created a dummy variable for *reform country*, encompassing countries that indicated in our surveys that CNA participation was motivated by the onset of a new education reform. Finally, we analogously coded *undetermined reform* for countries that participated in CNAs but for which we could not determine whether their initial participation was due to a reform.[60]

The results are presented in Table 3. If selection bias is an important problem, we would expect to see a large difference between reform and non-reform countries. There is some evidence

[58] We will discuss the survey in greater detail in the subsequent section. The specific question was: "How much did the following factors influence your country/economy's decision to participate in global international assessment for the first time? The onset of a new education reform." We separate countries that answered 4 ("Not at all") and 3 ("Very little") from those that answered 2 ("Somewhat") and 1 ("To a great extent").

[59] Countries that answered in the affirmative ("reform countries") are cases where the effects we are attributing to CNA participation may be in part due to coincidental education reform efforts. For countries that answered in the negative ("non-reform countries"), we can be more confident that we are observing the effect of CNA participation per se. The countries answering in the negative, i.e. that CNA participation was not due to education reform, were: Belgium, Botswana, Canada, Croatia, Denmark, Hungary, Iran, Korea, the Netherlands, Norway, New Zealand, Portugal, Sweden, Taiwan, Trinidad & Tobago, Tunisia, Ukraine, USA, and Yemen.

[60] These are countries that did not participate in our survey.

for this: the coefficient for reform countries is about twice as large as that for non-reform countries. This is intuitive: in countries where education reforms are ongoing when CNA participation is initiated, secondary enrollment is boosted not only from the effects of CNA participation, but also due to existing reform initiatives. However, more importantly for our purposes, the coefficient for non-reform countries is also positive and statistically significant. Even among countries where no domestic education reform was ongoing, CNA participation is clearly associated with increases in secondary enrollment.

We performed several additional robustness checks. We reran the analysis in Table 3 using an alternative measure collected by Braga et al, who identified major education reforms in 24 developed European countries from 1929-2000.[61] We coded non-reform countries as those that did not implement an education reform within the five years prior to CNA administration. This analysis produced analogous results: there was a positive association between CNA participation and higher net secondary enrollment even for countries with no prior reforms. As an alternative strategy to account for selection bias, we reran the analysis after separating countries for which CNA participation was mandated by an international organization (i.e., the country did not self-select into participation).[62] Again, first time participation was positively and significantly associated with increases in net secondary enrollment even for countries that participated due to an external mandate.

For developing countries in our sample, there is a potential concern that increases in secondary enrollment could be a statistical artifact: CNA administration may lead to better census

[61] Braga, Checchi, and Meschi 2013.

[62] This information was coded from our survey data using a procedure analogous to the reform variable. The relevant survey question was "How much did the following factors influence your country/economy's decision to participate in a global international assessment for the first time? -Mandated by organizations that our country/economy is a member of (e.g. OECD)."

data on students in the education system, boosting official enrollment numbers.[63] We thus reran the analysis limiting the sample to OECD countries, for which this is unlikely to be a concern. This produced similar substantive results. We also reran the analyses using two alternative dependent variables to avoid overreliance on a single measure, and these variables were associated with first-time participation as expected: the survival rate to the last year of primary school (positively associated with participation) and adolescents out of school as a share of secondary school aged children (negatively associated with participation). We also separated net secondary enrollment for male and female students and confirmed that the results hold for both sets of students. Interestingly, the substantive impact of participation is about twice as large for female students, which may reflect the consequences of socialization – assessment and aid agencies often place a high priority on inclusivity and gender equality in education – or attempts to boost CNA performance by targeting traditionally underserved students.

Transnational Pressure

In this subsection, we consider the role of transnational pressures surrounding CNAs. For aid-dependent countries, transnational pressure from the donor community is an important feature of their engagement with CNAs. CNAs can enhance a country's accountability and reputation among international donors or identify specific shortcomings to justify future aid programs. If so, we might expect CNA participation to be associated with increased foreign aid inflows.

We repeat the analysis from the prior section, substituting official development assistance to the education sector as the dependent variable.[64] As this is an aid inflow measure, there is no

[63] See related work by Lee and Zhang 2016.
[64] OECD DAC 2015, measured in millions of US$

variation in the dependent variable for countries that receive no aid. All other details remain the same. The results are presented in Table 4. The first column presents our baseline model. As predicted, CNA participation is strongly associated with increases in education aid. As with the previous section, we also reran the models with three-year leads and lags, and the results were broadly similar.

Figure 4 presents the results of our empirical model that includes specific leads and lags to examine trends in aid before and after the initiation of CNA participation. These results more clearly exhibit two distinct phases, with a jump in aid during the year of the assessment, followed by a dip and a more sustained increase thereafter. As we noted earlier, our expectation is that CNAs affect education policymaking in two phases, with elite interaction primarily taking place in the run up to the assessment, and consequences of score publication occurring with somewhat of a lag. Our findings are consistent with these stylized facts – aid agencies are often directly involved in the planning and implementation phase of CNAs, and we would expect to see aid disbursements concurrent with the administration of the CNA at t=0. However, there also appears to be an ex post increase in aid flows from t+3 onward. This more plausibly reflects responses by aid agencies to the publication of CNA results.

We also examined whether CNA participation has an impact on another potential source of transnational pressure: investment flows.[65] We did so by running analogous analyses replacing the dependent variable with foreign direct investment inflows as a share of GDP.[66] Unlike our results for net secondary enrollment and foreign aid, there was no statistically meaningful relationship between CNA participation and FDI inflows across specifications. We also recoded the independent variable to reflect CNA "outperformance" and "underperformance" based on the

[65] Also see Lee and Matanock 2016.
[66] Source: World Bank, World Development Indicators.

rationale that investors may respond to test scores that are particularly high or low compared to expectations.[67] This also produced no statistically meaningful results. This indicates that CNA participation does not directly affect aggregate FDI inflows, though it is possible that investor decisions are affected in some way not captured by our statistical models.

In sum, our results indicate that CNA participation is associated with meaningful increases in net secondary enrollment, our proxy for education quality and attainment, as well as foreign aid to education. Furthermore, these increases appear to be largely independent of self-selection by participating states. In the next section, we will examine the specific pathways of CNA influence over policy outcomes by leveraging an original survey of education officials.

Pathways of Influence: Analysis of Survey Data

One common challenge in international relations research is the difficulty of analyzing causal mechanisms. This is no different in the study of GPAs. Many of the proposed pathways of influence associated with assessment power are difficult to measure or quantify – e.g. reputational and status concerns – or difficult to observe – e.g. informal interaction between an assessment agency and a target state. In this section, we address this empirical challenge by utilizing a survey of elite education officials responsible for the planning and administration of CNAs in their countries.

Unlike many other GPAs, CNAs are conducted with the direct involvement of education officials in the target country. We were able to participate in two international conferences that

[67] Specifically, we coded CNA outperformers (underperformers) as countries for which CNA scores / GDP per capita were one standard deviation above (below) the global average, and reran the analysis separately entering first-time participation by over-, under-, and average performers as independent variables in the model. CNA scores are based on Angrist, Patrinos, and Sclotter 2013.

involved these officials from all countries participating in two major global assessments, TIMSS and PIRLS. Because the survey respondents are education policy elites directly responsible for the planning and administration of CNAs, they are intimately familiar with the context of CNAs and education policy in their countries. More importantly, they are important subjects in their own right, as they are the relevant policy elites hypothesized to serve as conduits of assessment power. In the exposition below, we will also draw on in-depth interviews conducted with 48 officials from target countries, assessment agencies, and donor agencies (Table 5).[68]

Survey Description

We collected data at two international conferences in 2011 and 2012 sponsored by the IEA, the organization responsible for TIMMS and several other CNAs.[69] The conferences provided us with direct access to country delegates who were involved in high-level education policy discussions on topics such as participation, administration, and usages of international assessments. The number of countries participating in these international conferences was at a historical peak in 2012 due to increasing participation and the simultaneous administration of two major CNAs, TIMSS and PIRLS. The conferences involved the participation of 150 delegates[70]

[68] Many interviews were conducted on-site at IEA conferences in Austria and Singapore. We contacted other subjects through snowball sampling and interviewed them via Skype (e.g. officers from Chile, Hong Kong, and with representatives from other international organizations such as OECD, UNESCO IIEP, and UNESCO). Interviews with relevant policy-makers knowledgeable about Vietnam's involvement in PISA were conducted in person in Hanoi, Vietnam in June 2011.

[69] We are omitting reference to the official titles of the conferences to protect respondent confidentiality.

[70] These delegates are representatives sent by their respective ministries or departments of education and responsible for the administration of cross-national assessments and post-implementation analysis through their respective ministries or departments of education.

from 67 countries[71] responsible for the administration and post-implementation analysis of CNAs in their countries (Table 6). We received responses from 77 delegates representing 46 countries.[72] Due to potential bias that could result from non-response, we checked to see if there were statistically significant differences in several characteristics between countries that responded to the survey and those that did not. We ran bivariate comparisons on several variables: GDP/capita, net secondary enrollment rate, polity score, and on global dataset on education quality in Reading, Mathematics and Science at the secondary level (Table 7). There were no statistically significant differences between the two subgroups.

This survey data is useful for three purposes. First, it sheds light on the mechanisms through which CNAs affect education policy across contexts. When administering assessments, are education officials concerned about rankings, technical transfers, international status, domestic political repercussions, or pressure from third parties such as international organizations? These are questions that are difficult to address with country-level data. Second, the survey and interviews shed light on the policy discourse and contestation surrounding CNAs across a large number of countries. Third, the analysis reinforces the plausibility of our central claim that participation in CNAs impacts education policy and outcomes. It would be troubling for our theory if most education officials indicated that CNAs have no meaningful effect on policy. On its own, the affirmation of an impact by survey respondents would not be decisive – e.g., education officials might have an inflated perception of the influence of CNAs. However, the findings in this section

[71] For the purposes of our analysis, we treat several "partner economies" as countries as they have substantially greater autonomy over policies within their jurisdictions than typical subnational governments. These are Chinese Taipei (Taiwan), Hong Kong, and the Palestinian Authority (West Bank and Gaza).
[72] There are more respondents than number of countries/economies because more than one delegate represented each country and every participant received an individual survey link.

complement the panel analysis by illustrating that education officials do in fact generally perceive a meaningful impact of assessments on education policy outcomes.

Cross-national Assessments and Education Reforms

We begin with a basic question: what is the perceived association between CNAs and domestic education policies? Our survey respondents are well positioned to answer this question, as they are policy elites not only responsible for planning and implementing CNAs, but also part of and intimately familiar with their countries' education policymaking establishment. We should note that the survey respondents are not advocates for assessments, but generally education bureaucrats on rotation within their ministries who happened to be working on assessments when we conducted the survey. The conference at which we conducted the survey was largely technical, focusing on how to administer assessments rather than discussing their broader impact. As such, we do not have a strong prima facie basis to believe that our respondents would exhibit a positive bias regarding the impact of assessments. However, these findings should be interpreted with appropriate caution, as they reflect the subjective assessments of policy officials.

The survey asked respondents to answer an open-ended question about how CNAs impacted their country's education policy. Several of the respondents answered that their country had only recently initiated participation in CNAs, and therefore it was too early to determine their impact. Omitting these responses, about 22% of respondents answered that CNAs had no impact on their country's education policies. 70% of respondents provided examples of how CNAs had affected the substance of their country's education policy or curriculum, and 7% mentioned other

impacts such as the development of greater national assessment capacity, increases in the education budget, or greater participation in other CNAs.

Many respondents drew specific, direct connections between CNAs and the substantive content of education reform in their countries. For example, a representative from Hungary noted that "PISA and PIRLS reading results led to the extension of reading and math skill-improving classes beyond grade 4, in grades 5 and 6. Now instead of subjects in some classes, schools can provide these classes similar to what they do in primary schools Grade 1-4."[73] A representative from New Zealand similarly noted that "The impact of the early cycles of TIMSS was quite significant as a driver for math educational policy, with the establishment of a math and science taskforce and then the numeracy strategy."[74] Similarly, in Malta, "the new science education strategy was based on the findings from TIMSS."[75] In Botswana, the 2003 TIMSS results led to "changes in the curriculum for [grades] 1 to 3… [and] introduction of a programme called SMASSE (Strengthening of Math and Science Programme in Education."[76] In Iran, "policy-makers changed the content of textbooks in Science and Reading….and adapted our curriculum to the framework of TIMSS and PIRLS."[77] In South Korea, "PISA 2006 results showed the drop of science literacy ranking in Korea. Due to this drop, educational policy of science education for helping female students was implemented."[78]

It is important to note that many of the policy changes mentioned by our respondents were reforms based on information first revealed by CNAs or which directly incorporated expertise

[73] Answer to survey question
[74] Answer to survey question
[75] Author interview with the Director of Assessments from Malta, Vienna, Austria, December 2011
[76] Author interview with the Director of Assessments and three education officers from Botswana, Singapore, March 2013
[77] Author interview with an education officer from Iran, in Singapore, March 2013
[78] Answer to survey question

gained by participation. The presence of these types of CNA-dependent reforms strengthens the case that CNAs likely play an important role in education policymaking: such reforms are highly unlikely to have occurred in the counterfactual case where countries had opted out of participation. Our findings reinforce existing, mostly qualitative work that identifies a link between CNA participation and education reforms.[79]

To provide additional support for our proposition that CNA participation tends to accelerate education reforms, we collected data on education reforms for all countries that initiated participation in CNAs after 1980. Education reforms are difficult to quantify systemically due to differences in policy contexts (e.g. legislation vs. administrative measures; centralized vs. decentralized education systems), differences in information availability (e.g. developed vs. developing countries), and subjective judgments about what actually constitutes an education reform. These caveats notwithstanding, we identified education reforms based on World Bank and UNESCO documents[80] and examined a 10-year window around each country's first-time participation in a CNA. Based on this data, countries implemented an average of 0.4 education reforms per year in the 5 years prior to CNA participation, and this accelerated to 0.8 education reforms per year in the 5 years after participation. We performed the same comparison using an independently collected dataset covering education reforms in developed European countries for

[79] Abdul-Hamid, Abu-Lebdeh, and Patrinos 2011; Takayama 2008; Addey 2015; Grek 2009; Breakspear 2012.

[80] For the sake of consistency, we used two primary sources for all countries: (i) World Bank Project Appraisal Documents, and (ii) UNESCO IBE National Reports/database. The World Bank is the largest multilateral aid agency that finances education projects around the world. Whenever the country develops a program, the World Bank publishes a Project Appraisal Document (PAD). The PAD describes the country context, strategic objectives in the education sector, and identifies education reforms. These documents are translated into English and are available online via their website. We also consulted UNESCO's IBE database for a more comprehensive data collection of reforms developed by both the individual countries as well as projects supported by international entities. Within the documents, we looked for keywords such as reforms, projects, and education goals under the subsection of each country's "Major Reforms in Education." The year in which the country participated in an international assessment was coded as post-assessment (omitting this year does not have a large bearing on the findings). This process was conducted for a list of countries for a time span of 10 years (-5 years pre and +5 years post first time participation in an international assessment).

a different time period (1929-2000),[81] and the results were analogous: there were 0.1 education reforms per year in the 5 years before first-time CNA participation, and this accelerated to 0.3 reforms per year in the 5 years after (in years prior to the 10-year window, the rate was also 0.1 education reforms per year).

Elite Politics

Our survey generally showed agreement with questions associated with elite politics mechanisms. Survey respondents were unanimous that a primary motivation for participating in international assessments is "to compare our education quality with other countries or economies." It is striking that no other survey item received unanimous support from respondents, including, "to measure and understand the current state of educational quality of my country/economy." The raison d'être of CNAs unquestionably stems from its comparative function, i.e. the ability of countries to place themselves within a global ranking of education quality. Our interviews tell a similar picture: officials repeatedly pointed to the comparative aspect as a core rationale for CNAs. An official from Jordan noted, "When you participate in international studies, you know whether you are performing well or worse compared to others."[82] 56% of respondents agreed that CNA participation "improves our reputation/status in the international community," even though our sample included a large number of developing countries that typically have low scores.

We asked survey respondents to identify the specific countries against which they compare their own scores. Broadly, respondents split into three groups: 1. those that tend to compare their scores against the highest performers (e.g. Finland, South Korea, Singapore); 2. those that situate

[81] Braga, Checchi, and Meschi 2013.
[82] Author interview with an education officer from Jordan, Austria, Vienna, December 2011

themselves within a regional or linguistic grouping (e.g. Arab countries, English-language countries), and 3. those that compare their results to countries at a similar level of economic development (e.g. developing countries, OECD countries). Interestingly, countries that place themselves in the first group are not necessarily those with the highest scores themselves. By providing clear, comparative rankings of educational performance, CNAs establish an international status hierarchy and compel countries to either accept their position within a plausible peer group or to aspire to higher status.

Survey respondents frequently noted that CNAs facilitate learning through interaction with assessment agencies and foreign counterparts. 84% of respondents agreed that CNA participation "improves our capacity to conduct and evaluate our own assessments," and 73% agree that participation "facilitates exchange of information between countries/economies." A Chilean official emphasized that participation in PISA involved sending a large delegation to OECD for consultations, information exchange, and training, "a very very impressive experience… it was very important to have progress in our capabilities."[83] A delegate from Trinidad and Tobago indicated that CNAs allowed them to "interact with IEA to learn about best practices…and to validate our own standard system."[84] In Botswana, the main benefit of CNA participation "… is to improve our research skills."[85] By participating in CNAs, countries are able to acquire valuable technical expertise to evaluate and improve the substance of their domestic education policy.

Several of our subjects went further and noted that CNA participation had a deeper impact on the discourse and norms surrounding education in their countries by shifting attention towards

[83] Author interview with a government officer at the Ministry of Education in Chile, videoconference, May 2013
[84] Author interview with an education officer from Trinidad and Tobago, Singapore, March 2013
[85] Author interview with the Director of Assessments and three education officers from Botswana, Austria, December 2011

the goals, metrics, and standards supported by international assessment agencies.[86] A representative from Kuwait commented that, "Once our country was exposed to international meetings that talked about international assessments, the revolution of education began. We started many awareness campaigns to also educate the mass[es]."[87] Similarly, a delegate from Honduras noted that CNAs have shaped a "big debate about what quality of education means" by focusing attention on "very valid and reliable instruments" promulgated by international assessment agencies, which tend to focus on skills such as math and reading.[88]

We also found evidence that political leaders view CNA participation as a mechanism to motivate education officials and improve standards. In Brazil, President Cardoso supported joining PISA despite predicting that his country would "come out at the bottom," because he saw participation as a mechanism to improve domestic education performance.[89] Similarly, in Vietnam, an important turning point was the personal involvement of then Deputy Prime Minister Nguyen Thiện Nhân, who staked his personal reputation on PISA participation. In 2011, Nguyen signed a letter confirming his country's participation in PISA despite the fact that "lower-level ministry staff was unsure of what participation in PISA [meant] for Vietnam."[90] Nguyen saw assessments as a way to signal educational and economic competitiveness as well as train and motivate education policymakers to maintain high standards in order to retain Vietnam's position in international rankings.[91]

[86] Martens and Niemann 2013.
[87] Author interview with an education officer from Kuwait, Singapore, June 2013
[88] Author interview with an education officer from Honduras, Vienna, December 2011
[89] Author interview with the Director of Education at OECD, teleconference, February 2012
[90] Author interview with an education officer at the World Bank, Hanoi, Vietnam, June 2011
[91] Author interview with an official in the Ministry of Education and Training in the Government of Vietnam, Hanoi, Vietnam, June 2011

In short, our survey results and interviews indicate that elite politics is a major source of influence for CNAs. The nature of CNAs, which require extensive interaction and consultation with assessment agencies, provides ample opportunities for learning, professionalization, and norm diffusion. By quantifying education quality and explicitly ranking countries, CNAs also trigger status and reputational concerns, creating an impetus for policy change.

Domestic Politics

Our survey respondents provided mixed views regarding the importance of domestic political considerations in the administration of CNAs. 62% of our respondents agreed that CNAs "improve our accountability with our citizens," and 43% agreed that "Negative [test] results could result in public upheaval." However, we note that our respondents, who were generally not political officials but career bureaucrats, may be relatively insulated from public pressures associated with CNAs. In addition, as we discuss below, our sample includes both democratic and autocratic states, and officials from the latter may be relatively less concerned about public pressures.

Several delegates we interviewed noted that their country's relative ranking can shape the domestic education policy discourse and accelerate reforms. For example, a US delegate noted that "TIMSS results and PISA results provided data to justify the sweeping reforms of the No Child Left Behind Act," though the impact was primarily "rhetorical justification" rather than shaping the content of the legislation.[92] Japan's declining rankings in international assessments built political momentum in favor of abandoning *yutori kyoiku*, which emphasized creativity under a relaxed curriculum, to one that emphasized more traditional rote and academic rigor.[93] Poor

[92] Answer to survey question
[93] Takayama 2008.

Hungarian performance in PISA since 2006 ignited protests, calls for the resignation of government ministers, and reforms such as the centralization of authority over the education system.[94]

Chile provides an informative case of how CNAs can alter the public discourse surrounding education and increase public accountability. After the fall of Pinochet's authoritarian regime, the newly established democratic government sought to differentiate itself by placing a strong emphasis on transparency, equity, and efficiency. As such, the government committed to participating in CNAs and making the results publically available. This sharply contrasted with the previous regime, which did not disclose results from their national assessments.[95] As a Chilean official noted, "in that context, to have an international, cross-national examination that… shows your relative position as these tests developed, announces to you and your population if you are advancing or not, or if you are going back in terms of results, it's of extreme importance."[96]

In Chile, CNAs provide a language and framework through which key stakeholders, such as teachers, parents, and students, articulate their concerns and frustrations. During the *Pingüino* Movement of 2006, college students criticized Chile's poor relative CNA performance compared to other OECD countries. Protesters expressed their concerns and used the results of CNAs to argue that the Chilean government must do more to improve the quality of education.[97] Chile's decision to drop out of TIMSS 2007 illustrates the potential salience of domestic political mechanisms. Although the Chilean government explained that it intended to focus on PISA and ongoing national assessments in lieu of TIMSS, it came under intense criticism for its decision:

[94] Blanka Zoldi, "Poor PISA Looms over Hungarian Education Reform," *IntelliNews*, 1-6-2017.
[95] Kijima and Leer 2015.
[96] Author interview with the Director of Evaluation and Curriculum at the Ministry of Education in Chile, videoconference, May 2013
[97] Kijima and Leer 2015.

"[the opposition party and media] accused the government of skipping TIMSS because the results of the reform were so poor that we were hiding... Politically, it was a disaster. From there onwards, there's no major international assessment that Chile skips."[98]

In sum, there is some basis to believe that the domestic political channel is an important source of assessment power for CNAs, particularly in democratic or democratizing states. The Chilean case demonstrates that parties newly in power may seek CNA participation as a means of demonstrating their commitment to transparency and reform. Teachers, students, and opposition politicians have used the results of CNAs to voice their discontent with the status quo. In effect, participation in CNAs has the potential to mobilize civic participation and intensify domestic demands for education policy reform.

Transnational Politics

According to our survey respondents, the role of transnational pressure as a pathway of influence for CNAs is somewhat limited. Only 22% of respondents agreed that CNA participation "attracts investors" to their country, and an even lower share of respondents, 12%, agreed that CNAs lead to "more resources/foreign aid to education by donors who credit our effort." However, it is important to note that our survey includes a large share of advanced industrialized countries, which may be less responsive to transnational economic pressure compared to developing countries.

Among developing countries, some officials perceive that CNA results lead to greater foreign investment. In Malaysia, there is a strong understanding among elites that investment in

[98] Author interview with the Director of Evaluation and Curriculum at the Ministry of Education in Chile, videoconference, May 2013

human capital is directly tied to economic growth. According to two officers at the Ministry of Education, "The Prime Minister goes to other countries to promote us. [But] investors went to China, because it's cheap labor. After the [TIMSS] test results, however, investors started coming back to Malaysia, because we have intellectual human capital…The test results directly impact investors' decisions."[99] Furthermore, assessments provide countries with an indicator of the quality of their human capital. In Botswana, assessments are used to "build a profile of competitiveness. These indexes are indicators of economic growth."[100] In some countries, the perception that CNA results are tied to foreign investment flows or economic growth appear to provide an additional motivation to implement quality improvements.

Among developing countries, multilateral and bilateral aid agencies are an important source of third-party influence. Donors occasionally step in to ease the financial burden on countries that require resources to implement CNAs. In Colombia, the Inter-American Development Bank provided supplementary funding to cover costs associated with analyzing TIMSS results.[101] The Palestinian Authority also receives support from the Finnish development aid agency to conduct technical analysis of its education system using TIMSS results.[102] While Kuwait does not receive any financing from the World Bank, it does receive technical assistance to improve its capacity to conduct domestic assessments and obtain information on international assessments.[103]

[99] Author interview with a Director of Assessment and an education officer from the Ministry of Education in Malaysia, Singapore, June, 2012
[100] Author interview with a Director of Assessment and three other education officers from the Ministry of Education in Botswana, Vienna, Austria, December, 2011
[101] Answer from survey question
[102] Author interview with the Director of Evaluation from West Bank and Gaza, Vienna, Austria, December, 2011
[103] Answer from survey question

Furthermore, aid agencies have a vested interest in the administration of CNAs in their client countries in order to monitor progress in the education sector. A government officer in Jordan indicated, "Most of our programs are funded by other agencies, like [the] World Bank, UNDP, Arab organizations, and EU. They want to have access to indicators. These organizations want some kind of indicators [to measure the progress] of the reform."[104] While donor agencies provide assistance to help improve education, the donor community is also "interested to see the results of their support for countries."[105] Hence, CNAs give donor agencies a point of leverage to influence national education policies, and the quantitative data necessary to monitor and evaluate the progress of education projects.

Cross-national Variation in Pathways of Influence

In the supporting information, we analyze our survey data using ordered probit models to examine how country-specific factors affect the mechanisms through which CNAs exert influence. The findings are largely consistent with our expectations: respondents in democratic countries are more likely to agree with survey questions associated with domestic political mechanisms, and respondents in aid-dependent countries are more likely to agree that receiving foreign aid is an important benefit of CNA participation.

However, the results also reveal some surprising patterns that may be worthy of further exploration. For example, we find that democracy and NGO density cut in the opposite direction in many of the models: policymakers in democratic states with low NGO density, such as Botswana, Ghana, Honduras, Lithuania, and Trinidad & Tobago, appear to be particularly

[104] Author interview with the Director of Evaluation from Jordan, Vienna, Austria, December, 2012
[105] Author interview with officers in charge of basic education from Yemen, Vienna, Austria, December, 2011

concerned with the status implications and domestic political repercussions of CNA participation. We also find that large exposure to inward FDI flows does not make respondents more likely to agree that CNA participation attracts private investors. This is broadly consistent with the findings from our panel analysis, which show that CNA participation is not associated with an increase in FDI inflows.

Conclusion

We have argued that cross-national assessments in education increasingly play an important role in the determination of education policy and outcomes. CNAs have proliferated rapidly over the past three decades, bringing comparability, transparency, and accountability to education policymaking. CNA participation increases the capacity of education officials and also provides incentives to facilitate improvements in education quality. We found strong support for our theory across all three sources of empirical evidence examined: a difference-in-differences analysis of panel data, survey data, and personal interviews with education policymaking elites. CNAs are a quintessential source of assessment power.

The nature of CNAs makes *elite politics* a particularly powerful pathway of influence. CNAs inherently involve close coordination between target states and authoritative organizations (i.e. assessment and donor agencies), providing clear opportunities for learning, professionalization, and norm diffusion. The comparative and transparent nature of CNAs also strongly evokes reputational and status concerns, motivating policymakers to improve education performance and climb international ranking tables. The findings from both our panel and survey evidence support the importance of elite mechanisms.

The observable impact of *domestic political pressures* may be somewhat limited for CNAs due to politicization of education reforms and self-selection, i.e. non-participation of leaders particularly concerned about domestic backlash. Nonetheless, we found qualitative evidence of mobilization around CNAs by citizens and political parties, particularly in democracies and democratizing states. Our survey also suggests that education officials in more democratic states tend to be more cognizant of the domestic political consequences of CNAs.

Our results suggest that *transnational pressure* is most relevant vis-à-vis aid-dependent developing countries, which are sensitive to the economic implications of CNA participation. Donor agencies see CNAs as a method to better evaluate education development projects and therefore improve education outcomes. Officials in some developing countries see CNAs as a way to signal high labor quality to potential investors. Our panel analysis indicates that CNA participation is associated with an increase in foreign aid inflows to education, but not FDI inflows.

The rapid growth in the number of CNAs and participants represents an important shift in global education policymaking. Assessment agencies and international organizations play an increasingly influential role in how countries discuss, design, and evaluate education policy. The rapid adoption of CNAs worldwide has coincided with the evolution of education from a national to a global issue[106] and an increasing recognition that education is a basic human right and global public good.[107] Although CNAs have not been without critics, this paper shows how participation can positively impact education outcomes. CNA participation enhances domestic and international accountability and provides policymakers with tools necessary to make informed decisions about the effective allocation of educational resources. Our findings link CNA participation to education reforms and increases in secondary enrollment and foreign aid inflows.

[106] Steiner-Khamsi 2003.
[107] Tsutsui and Wotipka 2004; Ramirez, Suárez, and Meyer 2007; Meyer, Bromley, and Ramirez 2010.

Nonetheless, we should not be entirely dismissive of critics who see CNAs as "educational colonialism" and an exertion of power by unaccountable, mostly Western technocrats.[108] Our findings show nontrivial contestation surrounding CNAs, such as concerns in developing countries that assessments do not adequately reflect cross-national variation in cultural, ethnic, or linguistic diversity. [109] In some cases, this may actually benefit students underserved by conventional approaches to education rooted in local culture and traditions – we find a particularly large substantive association between CNA participation and female enrollment. However, we are cognizant that our approach does not account for facets of education that are difficult to measure, such as ethical and personal development or artistic expression. We would caution that it is important for governments not to focus solely on rankings and test scores but to use CNAs as one of many inputs to improve the overall quality of education.

[108] Meyer 2014.

[109] 57% of developing country respondents agreed that "Test items and surveys do not accurately reflect our country/economy's cultural or ethnic diversity" compared to 14% for developed-country delegates. The percentages were identical to a similar question about linguistic diversity.

References

Abdul-Hamid, Husein, Khattab M Abu-Lebdeh, and Harry A Patrinos. 2011. Assessment testing can be used to inform policy decisions: the case of Jordan. *World Bank Policy Research Working Paper Series, Vol*.

Addey, Camilla. 2015. Participating in International Literacy Assessments in Lao PDR and Mongolia: A Global Ritual of Belonging. *Literacy as numbers: Researching the politics and practices of international literacy assessment*: 147–164.

Anderson, Benedict. 2006. *Imagined communities: Reflections on the origin and spread of nationalism*. Verso Books.

Angrist, Noam, Harry A Patrinos, and Martin Schlotter. 2013. An expansion of a global data set on educational quality: a focus on achievement in developing countries. *World Bank Policy Research Working Paper* (6536).

Artur, David. 2003. Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics, January*.

Baum, Matthew A, and David A Lake. 2003. The political economy of growth: democracy and human capital. *American Journal of Political Science* 47 (2): 333–347.

Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas, and Victor Pouliquen. 2013. *Turning a shove into a nudge? a 'labeled cash transfer' for education*. National Bureau of Economic Research.

Benveniste, Luis. 1999. *The Politics of Student Testing: a Comparative Analysis of National Assessment Systems in Southern Cone Countries.*

Bloom, David E. 2006. Measuring global educational progress. *Educating All Children: A Global Agenda*: 84–92.

Boulding, Carew. 2014. *NGOs, political protest, and civil society*. Cambridge University Press.

Braga, Michela, Daniele Checchi, and Elena Meschi. 2013. Educational policies in a long-run perspective. *Economic Policy* 28 (73): 45–100.

Breakspear, Simon. 2012. *The Policy Impact of PISA*. OECD Education Working Papers. Paris: Organisation for Economic Co-operation and Development. Available from <http://www.oecd-ilibrary.org/content/workingpaper/5k9fdfqffr28-en>. . Accessed14 April 2017.

Carnoy, Martin. 2014. Globalization, Educational Change, and the National State. In *Globalization and Education*. 2nd ed. R&L Education.

Chabbott, Colette. 2003. *Constructing educational development: international development organizations and the World Conference on Education for All*. Falmer Press.

Delaney, Jennifer A, and Patricia Yu. 2013. Policy innovation and tertiary education graduation rates: a cross-country analysis. *Compare: A Journal of Comparative and International Education* 43 (3): 387–409.

Dorius, Shawn F. 2013. The Rise and Fall of Worldwide Education Inequality from 1870 to 2010 Measurement and Trends. *Sociology of Education* 86 (2): 158–173.

Duflo, Esther, Rema Hanna, and Stephen P Ryan. 2012. Incentives work: Getting teachers to come to school. *The American Economic Review*: 1241–1278.

Greaney, Vincent, and Thomas Kellaghan. 2008. *Assessing national achievement levels in education*. Vol. 1. World Bank Publications.

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

Hanushek, Eric A, and Dennis D Kimko. 2000. Schooling, labor-force quality, and the growth of nations. *American economic review*: 1184–1208.
Hanushek, Eric A, and Ludger Woessmann. 2012. Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth* 17 (4): 267–321.
Hanushek, Eric A, and Ludger Woessmann. 2008. The role of cognitive skills in economic development. *Journal of Economic Literature* 46 (3): 607–668.
Henry, Miriam, Bob Lingard, Fazal Rizvi, and Sandra Taylor. 2001. *The OECD, Globalisation and Education Policy*. Published for IAU Press, Pergamon.
Honig, Dan. 2016. Seeing is Believing: The Normative Drivers of Agency Response to the Aid Transparency Index. Unpublished manuscript.
Husén, Torsten. 1979. An international research venture in retrospect: The IEA surveys. *Comparative Education Review* 23 (3): 371–385.
Jennings, Jennifer L, and Jonathan Marc Bearak. 2014. 'Teaching to the Test' in the NCLB Era How Test Predictability Affects Our Understanding of Student Performance. *Educational Researcher*: 0013189X14554449.
Jo, Hyeran, Bryan Phillips, and Joshua Alley. 2016. One Man's Terrorist: When Does Blacklisting Reduce Terrorism? Unpublished manuscript.
Kamens, David H, and Connie L McNeely. 2010. Globalization and the growth of international educational testing and national assessment. *Comparative Education Review* 54 (1): 5–25.
Keck, Margaret E, and Kathryn Sikkink. 1998. *Activists beyond borders: Advocacy networks in international politics*. Vol. 35. Cambridge University Press.
Kelley, Judith G., Beth A. Simmons, and Rush Doshi. 2016. The Power of Ranking: The East of Doing Business Indicator as a form of Social Pressure.
Kijima, Rie. 2013. The Politics of Cross-National Assessments: Global Trends and National Interests. Stanford University, Ph.D. Dissertation.
Kijima, Rie. 2010. Why Participate? Cross-National Assessments and Foreign Aid to Education. In *The Impact of International Achievement Studies on National Education Policymaking*, 13:. Vol. 13. International Perspectives on Education and Society Series. Emerald Group Publishing.
Kijima, Rie, and Jane Leer. 2015. Legitimacy, State-Building, and Contestation in Education Policy Development: Chile's Involvement in Cross-National Assessments. In *The Global Testing Culture: Shaping Education Policy, Perceptions, and Practice*. Oxford Studies in Comparative Education.
Koliev, Faradj, Thomas Sommerer, and Jonas Tallberg. 2016. Social Pressure and State Compliance. Unpublished manuscript.
Koretz, Daniel M. 2009. *Measuring up: what educational testing really tells us*. Harvard University Press.
Kremer, Michael, Conner Brannen, and Rachel Glennerster. 2013. The challenge of education and learning in the developing world. *Science* 340 (6130): 297–300.
Ladd, Helen F. 2002. School vouchers: A critical view. *The Journal of Economic Perspectives* 16 (4): 3–24.
Le, Anh, and Edmund Malesky. 2016. Do Subnational Governance Indices Lead to Improved Governance? Evidence from Field Experiment in Vietnam. Unpublished manuscript.

Lee, Melissa M., and Aila Matanock. 2016. Bottom Ten: Do State Failure and Fragility Indicators Affect Aid or Investment? Unpublished manuscript.
Lee, Melissa M., and Nan Zhang. 2016. Legibility and the Informational Foundations of State Capacity. *The Journal of Politics* 79 (1): 118–132.
Lockheed, Marlaine. 2012. Policies, performance and panaceas: the role of international large-scale assessments in developing countries. *Compare: A Journal of Comparative and International Education* 42 (3): 512–518.
Loveless, Tom. 2012. *Misinterpreting International Test Scores. In 2012 Brown Center Report on American Education.*
Martens, Kerstin, and Dennis Niemann. 2013. When Do Numbers Count? The Differential Impact of the PISA Rating and Ranking on Education Policy in Germany and the US. *German Politics* 22 (3): 314–332.
Meyer, Heinz-Dieter, and Aaron Benavot. 2013. PISA, power, and policy: The emergence of global educational governance. Symposium Books Ltd.
Meyer, John, Patricia Bromley, and Francisco Ramirez. 2010. Human Rights in Social Science Textbooks Cross-national Analyses, 1970–2008. *Sociology of Education* 83 (2): 111–134.
Miguel, Edward, and Michael Kremer. 2004. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 72 (1): 159–217.
Molutsi, Patrick P, and John D Holm. 1990. Developing democracy when civil society is weak: the case of Botswana. *African Affairs* 89 (356): 323–340.
Morse, Julia. 2016. Pathways to Policy Change: Blacklists and Market Enforcement in the Regime to Combat Terrorist Financing.
Mullis, Ina, and Michael Martin. 2007. TIMSS in Perspective: Lessons Learned from IEA's Four Decade of International Mathematics Assessments. In *Lessons Learned: What International Assessments Tell Us about Math Achievement.* International Perspectives on Education and Society Series. Brookings Institution Press.
Mundy, Karen. 1999. Educational multilateralism and world (dis) order. *Comparative Education Review* 42 (4): 448–478.
Mundy, Karen, and Caroline Manion. 2014. Globalization and Global Governance in Education. In *Globalization and Education*. 2nd ed.
Mundy, Karen, and Lynn Murphy. 2001. Transnational advocacy, global civil society? Emerging evidence from the field of education. *Comparative education review* 45 (1): 85–126.
Ramirez, Francisco O, John W. Meyer, and Julia Lerch. 2015. World Society and the Globalization of Educational Policy. In *The Handbook of Global Policy and Policy-Making in Education*.
Ramirez, Francisco O, David Suárez, and John W Meyer. 2007. The worldwide rise of human rights education. In *School knowledge in comparative and historical perspective*, 35–52. Springer.
Rodrik, Dani. 1995. Getting interventions right: how South Korea and Taiwan grew rich. *Economic Policy* 10 (20): 53–107.
Smith, William. 2016. Introduction. In *The Global Testing Culture: Shaping Education Policy, Perceptions, and Practice*. Oxford Studies in Comparative Education. Oxford: Symposium Books.
Steiner-Khamsi, Gita. 2003. The Politics of League Tables. *Journal of Social Science Education* 2 (1). Available from <http://www.jsse.org/index.php/jsse/article/view/470>. . Accessed14 April 2017.

Stone, Leslie F, Michelle Fryer, Monika Huppi, Ursula Quijano, Grace Noboa-Hidalgo, Patricia Sadeghi, Juliana Arbeláez, Magda Raupp, Miguel Székely, and Alejandro Cruz Fano. 2013. *Review of IDB Support to Secondary Education: Improving Access Quality and Institutions, 1995-2012*. Inter-American Development Bank.

Suárez, David F, Francisco O Ramirez, and Jeong-Woo Koo. 2009. UNESCO and the associated schools project: Symbolic affirmation of world community, international understanding, and human rights. *Sociology of Education* 82 (3): 197–216.

Sutton, Rosemary E. 2004. Teaching under High-Stakes Testing Dilemmas and Decisions of a Teacher Educator. *Journal of Teacher Education* 55 (5): 463–475.

Takayama, Keita. 2010. Politics of externalization in reflexive times: Reinventing Japanese education reform discourses through 'Finnish PISA success'. *Comparative Education Review* 54 (1): 51–75.

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

Tsutsui, Kiyoteru, and Christine Min Wotipka. 2004. Global civil society and the international human rights movement: Citizen participation in human rights international nongovernmental organizations. *Social Forces* 83 (2): 587–620.

UNESCO Institute for Statistics. 2010. *Global Education Digest 2010: Comparing Education Statistics Across the World*. UNESCO-UIS, Montreal, Quebec.

Vandenbussche, Jérôme, Philippe Aghion, and Costas Meghir. 2006. Growth, distance to frontier and composition of human capital. *Journal of Economic Growth* 11 (2): 97–127.

Volante, Louis. 2004. Teaching to the Test: What Every Educator and Policy-Maker Should Know. *Canadian Journal of Educational Administration and Policy*.

Weaver, Catherine. 2016. The Power and Politics of Aid Transparency Rankings and Ratings. Unpublished manuscript.

World Bank Group. 2011. Learning for All: Investing in People's Knowledge and Skills to Promote Development. The World Bank. Available from <http://siteresources.worldbank.org/EDUCATION/Resources/ESSU/Education_Strategy_4_12_2011.pdf>. . Accessed8 May 2013.

Figure 1: The Number of Participants in Cross-National Assessments in Education (1959-2012)

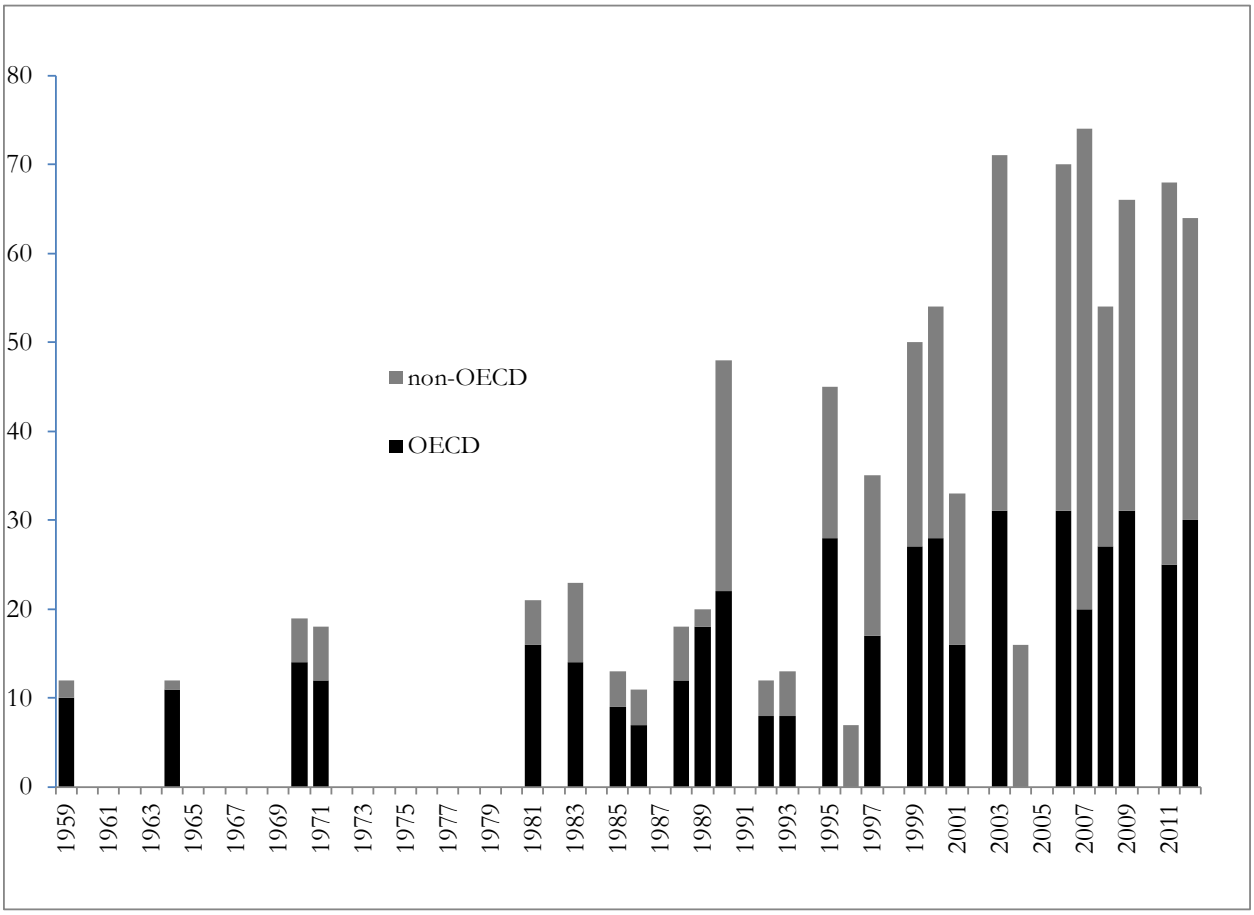
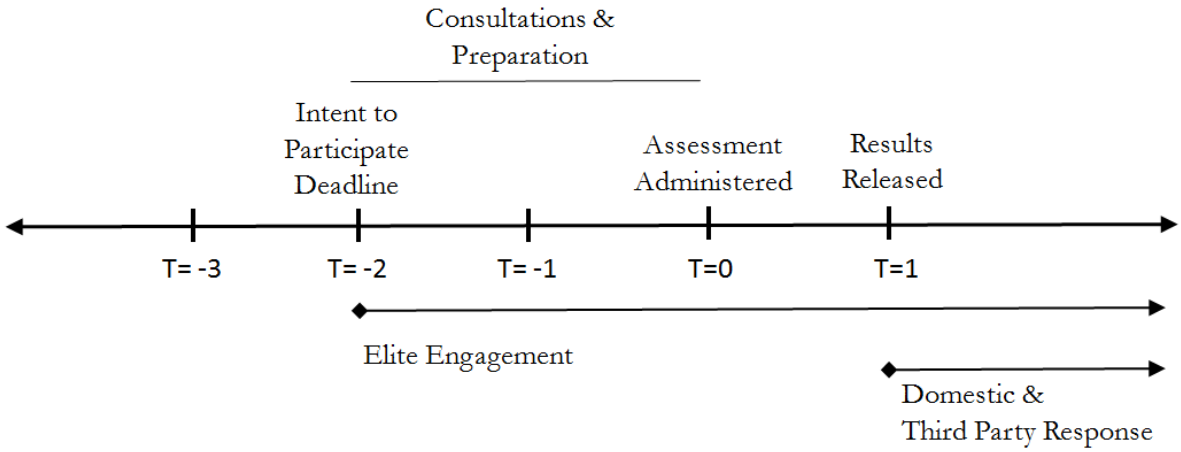


Figure 2: The Logistics of Participation in a CNA



Note: Each T is one year. The specifics of this timeline are based on the administration of PISA 2015, but the timeline is typical based on our review of 23 additional CNAs.

Figure 3: Estimated Impact of CNA Participation on Net Secondary Enrollment (%) for Years Before, During, and After Participation

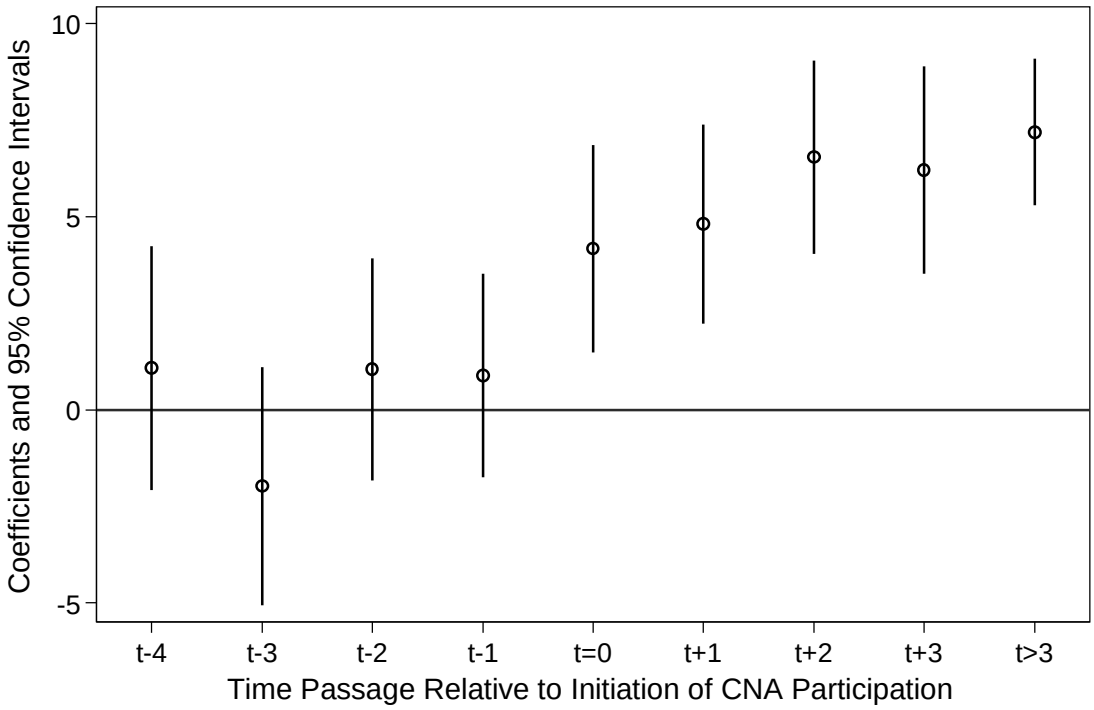


Figure 4: Estimated Impact of CNA Participation on Foreign Aid to Education ($US million) for Years Before, During, and After Participation

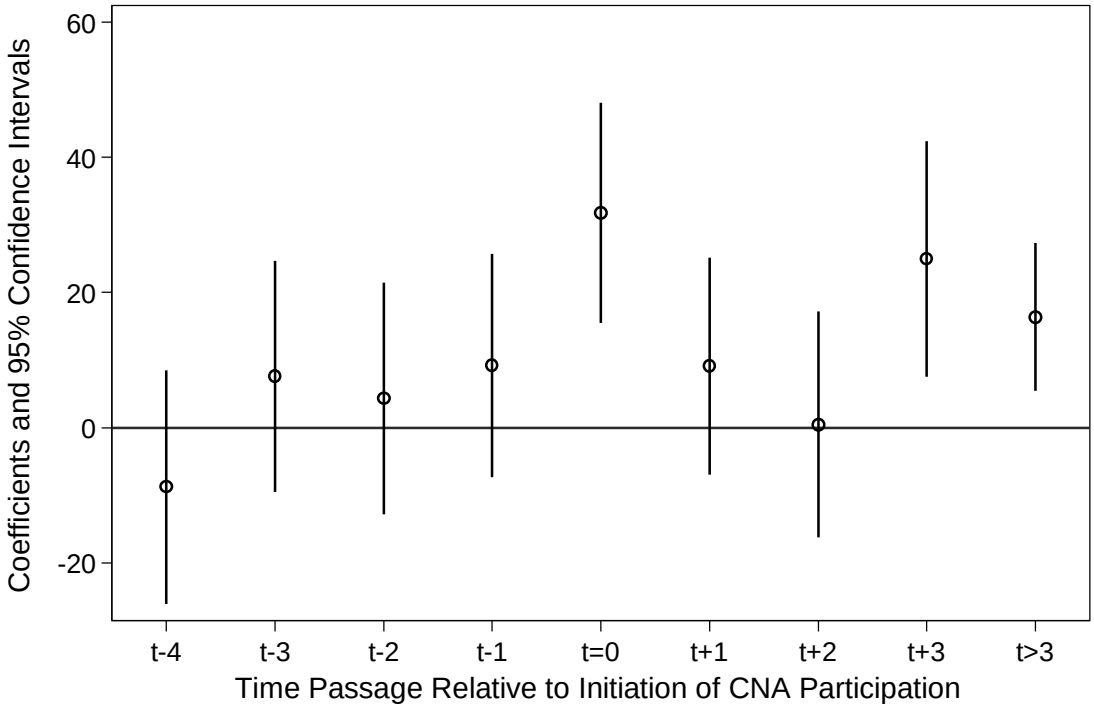


Table 1: List of Cross-National Assessments in the Panel Dataset

Name of Assessment	Year
Pilot Twelve-Country Study	1959
First International Mathematics Study	1964
First international Science Study	1970
Study of Literature Education	1970
Study of Reading Comprehension	1970
Study of Civic Education	1971
Study of English as a Foreign Language	1971
Study of French as a Foreign Language	1971
Classroom Environmental Study	1981
Second International Mathematics Study	1981
Second International Science Study	1983
Written Composition Study	1985
Pre-Primary Project	1986
International Assessment of Educational Progress	1988
Pre-Primary Project	1988
Computers in Education Study	1989
International Assessment of Educational Progress	1990
Reading Literacy Study	1990
Pacific Islands Literacy Level	1990
Computers in Education Study	1992
Pre-Primary Project	1993
Trends in Mathematics and Science Study	1995
Language Education Study	1995
South African Consortium for Monitoring Education Quality	1996
Latin American Educational Quality Assessment Laboratory	1997
International Civic and Citizenship Education Study	1997
Trends in Mathematics and Science Study	1999
International Civic and Citizenship Education Study	1999
Programme for International Student Assessment	2000
South African Consortium for Monitoring Education Quality	2000
Progress in International Reading Literacy Study	2001
Programme for International Student Assessment	2003
Trends in Mathematics and Science Study	2003
Second Regional Comparative and Explanatory Study	2004
Progress in International Reading Literacy Study	2006
Programme for the Analysis of the Educational Systems of Member Countries (CONFEMEN)	2006
Programme for International Student Assessment	2006

Pacific Islands Literacy Level	2006
South African Consortium for Monitoring Education Quality	2007
Trends in Mathematics and Science Study	2007
International Civic and Citizenship Education Study	2008
Teacher Education and Development Study in Mathematics	2008
Teaching and Learning International Survey	2008
Programme for the Analysis of the Educational Systems of Member Countries (CONFEMEN)	2008
International Civic and Citizenship Education Study	2009
Programme for International Student Assessment	2009
Progress in International reading literacy study	2011
Trends in Mathematics and Science Study	2011
Programme for International Student Assessment	2012

Table 2: Panel Analysis - Net Secondary Enrollment (OLS)

Indep/Dep Variables	Net Secondary Enrollment	Net Secondary Enrollment	Net Secondary Enrollment	Net Secondary Enrollment
Participation	6.61*	4.79*		
	(0.68)	(0.84)		
Participation (t-3)			5.09*	
			(0.65)	
Participation (t+3)				5.15*
				(0.74)
GDP/capita	-0.04	0.36*	-0.03	-0.02
	(0.03)	(0.07)	(0.03)	(0.03)
Democracy (Polity)	0.15*	-0.07	0.15*	0.12
	(0.06)	(0.08)	(0.06)	(0.07)
Education Aid		0.000		
		(0.003)		
Country Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y
n	1673	895	1673	1520

Note: Numbers in parenthesis are standard errors. Star denotes a coefficient at least two standard errors removed from zero.

Table 3: Panel Analysis – Separating Countries by the Onset of Domestic Reform (OLS)

Indep/Dep Variables	Secondary Enrollment
Non-Reform	3.62*
	(1.21)
Reform	8.22*
	(1.44)
Undetermined Reform	4.09*
	(0.76)
GDP/Capita	-0.04*
	(0.03)
Democracy (Polity)	0.19*
	(0.06)
Country Fixed Effects	Y
Year Fixed Effects	Y
n	1673

Note: Numbers in parenthesis are standard errors. Star denotes a coefficient at least two standard errors removed from zero.

Table 4: Panel Analysis - Education Aid (OLS)

Indep/Dep Variables	Education Aid	Education Aid	Education Aid
Participation	14.27* (4.06)		
Participation (t-3)		10.44* (4.09)	
Participation (t+3)			14.29* (4.10)
GDP/capita	-0.51 (0.45)	-0.45 (0.45)	-0.53 (0.46)
Democracy (Polity)	-0.02 (0.32)	-0.01 (0.32)	-0.03 (0.32)
Country Fixed Effects	Y	Y	Y
Year Fixed Effects	Y	Y	Y
n	3180	3180	3180

Note: Numbers in parenthesis are standard errors. Star denotes a coefficient at least two standard errors removed from zero. Education aid is measured in $US million.

Table 5: List of Personal Interviews

	Country/Org.	Region	Status	Date
1	Morocco	MENA	non-OECD	12/10/2011
2	IIEP		Agency	2/13/2012
3	Malaysia	EAP	non-OECD	12/8/2011
4	Indonesia	EAP	non-OECD	6/24/2012
5	Palestine	MENA	non-OECD	12/9/2011
6	Tunisia	MENA	non-OECD	12/6/2011
7	Taiwan	EAP	non-OECD	12/6/2011
8	Kazakhstan	ECA	non-OECD	6/24/2012
9	Yemen	MENA	non-OECD	6/25/2012
10	Thailand	EAP	non-OECD	6/22/2012
11	Vietnam	EAP	non-OECD	6/15/2011
12	Honduras	LAC	non-OECD	12/8/2011
13	South Africa	SSA	non-OECD	12/8/2011
14	Hong Kong	EAP	non-OECD	2/24/2012
15	Consultant		Agency	5/2/2011
16	Kazakhstan	ECA	non-OECD	12/6/2011
17	Taiwan	EAP	non-OECD	12/6/2011
18	IEA		Agency	5/4/2011
19	Iran	MENA	non-OECD	6/23/2012
20	Botswana	SSA	non-OECD	12/7/2011
21	Ireland	OECD	non-OECD	12/6/2011
22	Malta	ECA	non-OECD	12/6/2011
23	Kuwait	MENA	non-OECD	6/25/2012
24	Hong Kong	EAP	non-OECD	2/17/2012
25	OECD		Agency	2/14/2012
26	Consultant		Agency	2/1/2012
27	Vietnam	EAP	non-OECD	6/14/2011
28	Ghana	AFR	non-OECD	6/21/2012
29	Vietnam	EAP	non-OECD	6/13/2011
30	Russia	ECA	non-OECD	12/5/2011
31	Trinidad and Tobago	LAC	non-OECD	12/5/2011
32	Malaysia	EAP	non-OECD	6/22/2012
33	UAE, Dubai	MENA	non-OECD	12/7/2011
34	UAE	MENA	non-OECD	12/7/2011
45	Yemen	MENA	non-OECD	12/8/2011
46	Vietnam	EAP	non-OECD	6/15/2011
47	OECD		Agency	2/15/2012
48	Colombia	LAC	non-OECD	12/6/2011
49	Jordan	MENA	non-OECD	12/10/2011

40	South Africa	SSA	non-OECD	12/9/2011
41	Chile	LAC	OECD	12/8/2011
42	Vietnam	EAP	non-OECD	6/15/2011
43	UNESCO		Agency	2/14/2012
44	Chile	LAC	OECD	3/7/2013
45	Mexico	LAC	OECD	3/18/2013
46	Chile	LAC	OECD	5/3/2013
47	Chile	LAC	OECD	5/3/2013
48	Chile	LAC	OECD	5/3/2013

Table 6: List of Countries Surveyed (Responses received from bold & italicized)

OECD Countries	Non-OECD
Australia	***Armenia***
Austria	Azerbaijan
Belgium	Bahrain
Canada	***Botswana***
Czech Republic	***Bulgaria***
Denmark	***Chile***
Finland	***Colombia***
France	***Croatia***
Germany	Georgia
Hungary	***Ghana***
Ireland	***Honduras***
Israel	***Hong Kong SAR***
Italy	Indonesia
Japan	***Iran***
Netherlands	***Jordan***
New Zealand	***Kazakhstan***
Norway	***Kuwait***
Poland	Lebanon
Portugal	***Lithuania***
Slovakia	Macedonia, FYR
Slovenia	***Malaysia***
South Korea	Malta
Spain	Morocco
Sweden	***Oman***
Turkey	***Qatar***
United Kingdom	***Romania***
United States	***Russian Federation***
	Saudi Arabia
	Serbia
	Singapore
	South Africa
	Syrian Arab Republic
	Taiwan
	Thailand
	Trinidad and Tobago
	Tunisia
	UAE
	Ukraine
	West Bank and Gaza
	Yemen

Table 7: Comparison of Mean between Response and Non-Response Countries

	mean	standard error	t	df	P (different from zero)
GDP per capita					
Countries that did not respond	23,565	3,044	-0.42	61	0.68
Countries that responded	25,388	2,671			
Net Secondary Enrollment Rate					
Countries that did not respond	84.32	3.90	-0.73	36	0.47
Countries that responded	86.77	2.61			
Polity2 Score					
Countries that did not respond	4.74	1.57	-0.43	59	0.67
Countries that responded	5.52	0.99			
Educational Quality - Secondary					
Countries that did not respond	50.89	1.33	0.02	39	0.98
Countries that responded	50.85	1.07			
Educational Quality – Secondary, Mathematics					
Countries that did not respond	55.3	1.50	1.33	39	0.81
Countries that responded	54.7	1.30			
Educational Quality – Secondary, Reading					
Countries that did not respond	49.9	1.30	-0.02	39	0.98
Countries that responded	50.0	0.90			
Educational Quality – Secondary, Science					
Countries that did not respond	47.4	1.35	-0.26	38	0.79
Countries that responded	47.9	1.00			

Sources: WB GDP per Capita (2010), WDI Net Secondary Enrollment Rate (2010), Correlates of War, Polity2 IV score (2010), Education Quality Indicators (Angrist, Patrinos, and Schlotter, 2013) for assessments conducted in 2010.

SUPPORTING INFORMATION

Ordered Probit Analysis of Survey Data

In this supporting information section, we consider variation in the pathways of influence associated with CNAs using ordered probit models. CNAs will not necessarily impact policymaking in the same way in Finland and Kazakhstan and Botswana. In this section, we will consider how country-specific factors might affect the mechanisms through which CNAs exert influence.

In order to examine variation in causal mechanisms, we break out questions associated with specific pathways of influence from our survey. Using responses to these questions as dependent variables, we examine what country-specific covariates affect the likelihood of respondents answering in the affirmative. Specifically, for *Elite Response*, we use responses related to questions about reputation/status in the international community (ER1), leadership in the world (ER2), risk perceptions about low rankings (ER3), concern about the participation of economic competitors (ER4), and desire for technical support and expertise (ER5).[110] For *Domestic Politics*, we use responses related to questions about the potential for domestic public upheaval from poor performance (DP1), improving accountability with citizens (DP2), and citizen pressure (DP3).[111]

[110] The specific survey questions were: ER1: "What are the benefits associated with your country/economy's participation in global international assessments like PISA, TIMSS and/or PIRLS? Improves our reputation/status in the international community." ER2: "What are the main reasons why your country/economy participates in cross-national assessments? To be a leader in the world." ER3: "What are the main challenges associated with your country/economy's participation in global international assessments like PISA, TIMSS and/or PIRLS? Ranking low on the international scale is a risk." ER4: "What are the main reasons why your country/economy participates in cross-national assessments? Because our economic competitors are participating." ER5: "What are the main reasons why your country/economy participates in cross-national assessments? To obtain technical support and expertise to improve our own capacity for conducting assessments."

[111] The specific questions were: DP1: "What are the main challenges associated with your country/economy's participation in global international assessments like PISA, TIMSS and/or PIRLS? Negative results could result in public upheaval." DP2: "What are the benefits associated with your country/economy's participation in global international assessments like PISA, TIMSS and/or PIRLS? Improves our accountability with our citizens." DP3:

For *Transnational Pressure*, we use responses related to questions about CNAs as attracting investors (TP1), CNA participation in response to requests from donor agencies (TP2), and CNAs as a condition for aid disbursement (TP3).[112] Each of these survey questions touches on a somewhat different causal mechanism within the broader pathways of elite response, domestic politics, and transnational pressure.

We use independent variables that proxy for factors that might make countries particularly susceptible to influence through distinct mechanisms: 1. GDP per capita (PPP) is included to account for varying levels of economic development. The scope for policy change through learning and interaction with international assessment agencies is presumably greater for economically less-developed countries that lack the resources and knowhow of more developed peers. Less developed countries may also be more susceptible to transnational pressure from markets and aid agencies; 2. Net secondary enrollment (%) is a proxy for the quality of the domestic education system, as we discussed in the panel section;[113] 3. Gender Equality: because CNAs typically break out performance according to the student gender, officials from countries with large gender disparities in education may be particularly sensitive to reputational consequences. We use the Global Gender Gap Index, in which high scores indicate higher levels of gender equality; 4. Polity score: ceteris paribus, officials of more democratic countries are likely to be more receptive to domestic pathways of influence such as demands from civil society

"What are the main reasons why your country/economy participates in cross-national assessments? Pressures from citizens about showing results in the education sector."

[112] The specific questions were: TP1: "What are the benefits associated with your country/economy's participation in global international assessments like PISA, TIMSS and/or PIRLS? Attracts Investors." TP2: "What are the main reasons why your country/economy participates in cross-national assessments? Requests from donor agencies." TP3: "What are the main reasons why your country/economy participates in cross-national assessments? Participation in assessment is a conditionality for aid disbursement"

[113] We also tried substituting actual test scores in lieu of secondary enrollment. As several countries in our survey did not yet have test scores available, this results in fewer observations and less precision in our estimates. However, the substantive results were generally similar using test scores.

and citizens; 5. Country membership in international nongovernmental organizations (NGOs): a high concentration of NGOs may also increase the salience of domestic pathways of influence. As data on general NGO density is not readily available,[114] we use membership in international human rights NGOs. This is a somewhat crude proxy, as human rights NGOs are often concerned about a verity of issues aside from the right to education. However, the number of such NGOs is a plausible proxy for the strength of domestic civil society and the potential for NGOs to shape the domestic policymaking process[115]; 6. Intergovernmental organization (IGO) memberships: governments with dense ties to IGOs may be more open to or susceptible to external influence or pressure. Some IGOs mandate CNAs as a condition for membership or aid. Even where this is not the case, IGOs are often directly involved in the implementation and analysis of CNAs; 7. Inward Foreign Direct Investment / GDP: countries that receive large inward FDI flows may be more responsive to the consequences of CNAs on investor behavior. Participating in CNAs and performing well can send a signal to potential investors about the high quality of a country's labor force, facilitating greater investment flows; 8. Net official development assistance received per capita: dependence on foreign aid will likely make countries more susceptible to pressure by international aid agencies and bilateral donors.

As all of our dependent variables are measured as scales, with high numbers indicating agreement with the question and low numbers indicating disagreement, we use ordered probit for all specifications. Standard errors are clustered by country to account for multiple responses received from delegates representing the same country.

[114] Boulding 2014, 16.
[115] Suárez, Ramirez, and Koo 2009; Mundy and Murphy 2001; Keck and Sikkink 1998, vol. 35.

The results associated with Elite Politics are presented in Table 8. The dependent variable in column 1 is respondent agreement with the notion that CNAs improve their country's reputation/status in the international community. The results suggest that representatives of relatively democratic states with low NGO density and high aid dependence view CNAs as enhancing their reputations. The substantive effects associated with these variables are quite large. For example, holding all other variables to their mean values, the predicted probability of a delegate from an autocratic country with the minimum polity score of -10 (e.g. Qatar) agreeing or strongly agreeing with the reputation question was respectively 0.13 and 0.02,[116] while the same for a democratic country with the maximum polity score of 10 (e.g. Norway) was respectively 0.46 and 0.33.[117] The tendency for democracy and NGO density to cut in opposite directions is interesting, and we will return to this point below.

Column 2 indicates that representatives of economically developed countries are more likely to agree with the idea that CNAs allow their country to be a "world leader," suggesting that distinct reputational mechanisms may at be at work for different types of countries. Column 3 indicates that developed countries are less likely to view low rankings as a risk. Interestingly, our proxy for education quality – secondary school enrollment – is not a meaningful predictor of responses to these questions or any of the others in this table. Furthermore, we tried substituting actual CNA scores in lieu of secondary enrollment, and the results were substantially similar. Absolute CNA test scores per se do not appear to play a meaningful role in how education officials view their own countries' participation in CNAs. There are several reasons why this might be the case. For one, depending on their policy views, education officials may view low rankings as a net positive, for example if they lead to education reforms or larger education budgets. It may also be

[116] 95% confidence intervals respectively [0.00,0.45] and [0.00,0.18]
[117] 95% confidence intervals respectively [0.30,0.62] and [0.15,0.57]

that low-ranking countries are able to justify their weak absolute performance by comparing themselves to other low-ranking countries in their region or other peer group.

As one might expect, column 4 shows that delegates were more likely to agree that the status of their economic competitors matters in their decision-making about CNAs if their country is a relatively large recipient of inward FDI and foreign aid. High NGO density also appears to be associated with concern about economic competitors. Column 5 indicates that relatively poor countries with low NGO density tend to see the technical support and expertise obtained from participation in CNAs as an important benefit of participation. These are countries like Ghana and Yemen, which do not have strong domestic capacity in education.

Turning to Table 9, across all three of the dependent variables, more democratic states tend to exhibit greater concern regarding the domestic political implications of CNAs. This is consistent with our theoretical predictions, though NGO density again generally enters with a negative sign. Countries with stronger gender equality tend to be less concerned about public upheaval associated with CNA results and are less likely to cite public accountability as a reason for participation.

Table 10 presents results associated with transnational politics. As predicted, high aid-dependence is associated with respondents agreeing that donor requests and conditionality are important. IGO membership is also associated with affirmative responses for donor conditionality. The results for “attracting investors” are more puzzling. Inward FDI is not a meaningful predictor of agreement with this question. Instead, expressed concern with attracting investors is higher in democratic countries with low NGO density, low secondary school enrollment, and low gender equality.

The tendency for democracy and NGO density to cut in opposite directions in several of our empirical models is somewhat puzzling. In our dataset, there are several democratic states with

low NGO density, such as Botswana, Ghana, Honduras, Lithuania, and Trinidad & Tobago. Democratic governments with weak civil society face some distinct challenges of governance that may be contributing to these results. In particular, an active civil society tends to aggregate information about citizen preferences and facilitate predictable patterns of political interaction.[118] In the absence of a strong civil society, democratic leaders may be more concerned about valence politics, and poor CNA performance has the potential to raise questions about the competence of political leaders.

[118] Molutsi and Holm 1990.

Table 8: Survey Analysis - Elite Response (Ordered Probit)

Indep/Dep Variables	Improves Reputation/ Status	To be a World Leader	Low Ranking is a Risk	Economic Competitors Participate	Obtain Technical Support & Expertise
GDP/capita	0.009	0.012*	-0.012*	0.006	-0.015*
	(0.008)	(0.005)	(0.003)	(0.004)	(0.003)
Secondary Enrollment	0.032	0.011	-0.017	0.008	0.006
	(0.019)	(0.010)	(0.011)	(0.014)	(0.013)
Gender Equality	-3.079	-1.968	2.073	-1.540	1.705
	(4.458)	(3.086)	(2.883)	(3.175)	(3.200)
Democracy (Polity)	0.108*	0.058	-0.013	-0.002	0.004
	(0.051)	(0.034)	(0.022)	(0.029)	(0.028)
NGO	-1.491*	-0.249	-0.156	0.853*	-1.312*
	(0.447)	(0.576)	(0.304)	(0.374)	(0.406)
IGO	0.014	-0.018	-0.013	-0.025	0.033
	(0.017)	(0.020)	(0.015)	(0.015)	(0.019)
FDI/GDP	-0.053	-0.011	0.016	0.039*	0.013
	(0.030)	(0.025)	(0.014)	(0.020)	(0.015)
ODA/capita	0.006*	0.002	0.002	0.005*	-0.003
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
n	53	67	53	67	67

Note: Numbers in parenthesis are country-clustered standard errors. Star denotes a coefficient at least two standard errors removed from zero.

Table 9: Survey Analysis - Domestic Politics (Ordered Probit)

Indep/Dep Variables	Fear of Public Upheaval	Accountability With Citizens	Pressure From Citizens
GDP/capita	0.001	0.009	0.016*
	(0.007)	(0.006)	(0.007)
Secondary Enrollment	-0.006	0.008	-0.012
	(0.012)	(0.015)	(0.016)
Gender Equality	-9.902*	-9.868*	-6.170
	(2.651)	(4.330)	(4.605)
Democracy (Polity)	0.127*	0.106*	0.131*
	(0.035)	(0.041)	(0.045)
NGO	-0.894*	-0.587	-0.701
	(0.378)	(0.520)	(0.399)
IGO	0.039*	0.006	0.026
	(0.018)	(0.018)	(0.019)
FDI/GDP	0.037	-0.048	-0.060
	(0.022)	(0.027)	(0.049)
ODA/capita	0.001	-0.003	0.004
	(0.003)	(0.002)	(0.003)
n	53	53	56

Note: Numbers in parenthesis are country-clustered standard errors. Star denotes a coefficient at least two standard errors removed from zero.

Table 10: Survey Analysis - Transnational Pressure (Ordered Probit)

Indep/Dep Variables	Attracting Investors	Donor Agency Request	Aid Conditionality
GDP/capita	0.007	-0.029	-0.028
	(0.005)	(0.028)	(0.015)
Secondary Enrollment	-0.029*	0.007	-0.015
	(0.013)	(0.025)	(0.024)
Gender Equality	-14.454*	-9.064*	-7.204
	(4.047)	(6.494)	(5.406)
Democracy (Polity)	0.128*	0.314*	-0.058
	(0.030)	(0.153)	(0.061)
NGO	-1.122*	0.079	0.669
	(0.460)	(0.505)	(0.659)
IGO	0.046	0.036	0.054*
	(0.025)	(0.023)	(0.024)
FDI/GDP	0.027	-0.195*	-0.052
	(0.038)	(0.073)	(0.035)
ODA/capita	0.001	0.031*	0.012*
	(0.002)	(0.010)	(0.005)
n	53	53	56

Note: Numbers in parenthesis are country-clustered standard errors. Star denotes a coefficient at least two standard errors removed from zero.