

Business as Usual or Donor Politics? An Analysis of Peer Review Assignments in the OECD Development Assistance Committee

Simone Dietrich*

Alice Iannantuoni[†]

Bernhard Reinsberg[‡]

September 2021[§]

Abstract

The OECD Development Assistance Committee uses a peer review mechanism to monitor and evaluate its member states' foreign aid programs. While, to date, studies of peer reviews have focused largely on questions related to the effectiveness of peer reviews as a mechanism for disseminating best practices, we examine crucial role assignment decisions that form the foundation of the peer review process and that could either be shaped by a bureaucratic process or by the interest of individual donors. In particular, we focus on (i) the assignment of peer examiners to a given review, and (ii) the selection of recipient countries for in-depth field visits when the review process is ongoing. In both cases, we expect the OECD bureaucracy and individual donors to hold competing preferences as to which examiners or recipient countries get assigned. Our preliminary analyses show that the assignment of peer examiners is largely driven by the bureaucratic process, while the selection of recipient countries for field visits is shaped by the preferences of reviewed donors.

*Department of Political Science and International Relations, University of Geneva, simone.dietrich@unige.ch.

[†]Department of Political Science and International Relations, University of Geneva, alice.iannantuoni@unige.ch.

[‡]School of Social and Political Sciences, University of Glasgow, bernhard.reinsberg@glasgow.ac.uk.

[§]Preliminary draft, please do not share without the authors' permission. We wish to thank Katharina Fleiner and Tomas Spahini for research assistance as well as participants of the ECPR Joint Sessions Workshop "The Performance of Global Governance Institutions" for helpful comments on a previous draft.

1 Introduction

Increasing interdependence has led to the creation of many different global governance regimes, such as environment, finance, security, and trade (Alter and Meunier 2009; Keohane and Nye 1977; Raustiala and Victor 2004). What they share is the aim to address transnational problems. In this context, intergovernmental organizations have quickly moved beyond their initial functions of coordination to become global regulators producing rules that influence domestic governance and policy outcomes (Abbott and Snidal 2013; Farrell and Newman 2014; Prakash and Potoski 2016).

Intergovernmental organizations (IGOs)—as the most centralized actors in world politics beyond states—have long used hierarchical modes of governance. According to principal-agent scholarship, states delegate decision-making authority to IGOs, which perform certain functions on their behalf, such as monitoring state behaviors, enforcing policy decisions, and settling disputes among member states (Abbott and Snidal 1998; Brown 2010; Nielson and Tierney 2003). However, IGOs as global regulators are increasingly under siege. Not only are they ailing as platforms for action (Hale, Held, and Young 2013), but also they are challenged for lack of accountability and democratic legitimacy (Dingwerth, Schmidtke, and Weise 2011; Freyburg et al. 2017; Moravcsik 2004).

At the same time, states have developed non-hierarchical governance modalities structured around more limited delegation to IGO bureaucracies (Hooghe and Marks 2015; Vabulas and Snidal 2013; Westerwinter 2019). One such mechanism is peer review: the systematic examination of the performance of a state by other states, with the ultimate goal of helping the reviewed state improve its policy making, adopt best practices, and comply with established standards (Pagani 2002). Peer reviews are soft governance instruments that emerged in the mid-1990s and have since been proliferating (Albaret 2016; Conzelmann 2014; Liverani and Lundgren 2007). As instruments that help monitor and promote state compliance with international norms and regulations, peer reviews promote horizontal relations in global governance, where relations among states are characterized as collegial and mutually respecting their sovereignty. The ultimate goal of peer review is horizontal policy transfer, eventually leading to convergence toward a collectively shared policy consensus (Groenendijk 2012).

Our empirical case of peer review is the Organization for Economic Cooperation (OECD), specifically its Donor Assistance Committee (DAC) in which all major Western donors convene to develop norms of “good donorship.” Endowed with limited autonomous powers, the OECD Secretariat plays a facilitating role across a range of areas of global economic governance, fostering consensus among member states by providing a forum for dialogue and collective policy-making (Porter and Webb 2007). A body of previous work has recognized the role of OECD DAC peer review for the advancement of policy norms (Ashoff 2013;

Eyben 2013; Paulo and Reisen 2010), but analyses of the peer review mechanism have thus far focused on aid evaluation practices (Liverani and Lundgren 2007), education policy (Niemann and Martens 2018), tax policy (Webb 2004), and labor market and social policy (Mahon and McBride 2009). Recently, Carcelli (2021) studies how one particular OECD agreement, the 2001 DAC Recommendation on Untying Official Development Assistance, has shaped donor practices in tying foreign aid.

While, in a larger project, we are interested in understanding whether and how the OECD DAC shapes donor practices, this paper tackles an important foundational question that asks to what extent politics infiltrates key assignment choices upon which the architecture and the OECD DAC peer review process rests: (i) the choice of examiner countries, and (ii) the choice of aid recipient countries where the reviewed member's development assistance program and practices are assessed in-depth. Because the peer review is carried out by designated member states and the OECD Secretariat, there is a real possibility that politics influences these peer review assignment mechanisms—which, in turn may then impact the content and quality of a review.

In a first instance, we look at the process of assigning examiner countries, where donors are asked to assess aid practices of the DAC member (or reviewee) that is considered under review. As indicated by OECD DAC guidelines, the assignment of peer review teams should merely be a function of bureaucratic considerations and that donors under review should be paired with examiner countries that facilitate review, including, for example, shared language or a similar aid bureaucratic structure. However, it is also possible that this assignment mechanism is vulnerable to politics and donor interests. It could be, for example, that donors that are about to undergo review lobby the OECD Secretariat to be paired with examiners who share similar partisan orientations. Or, it could be that potential examiner countries lobby the OECD Secretariat to examine countries with whom their overlap in the foreign policy preferences. While the former rationale for assignment results from the bureaucratic process of the OECD Secretariat, the latter rationale suggests that politics may be at work and a deciding factor in the examiner assignment.

In a second instance, we look at the process of selecting recipient countries for an in-depth examination of a donor aid program. Over the course of the review, examiners travel to a small number of designated recipient countries to study donor practices. This choice of recipient countries could be a function of bureaucratic considerations insofar as they may be more likely to take place in countries that have a good evaluation infrastructure or that are among the most relevant for the OECD donor community as a whole. Alternatively, the choice of recipients may result from donor interest considerations where DAC member states that are being reviewed select recipients that are important to them or with whom they have long-standing relationships. Again, there are two competing explanatory models: one that captures factors relating to the bureaucratic process and the overall organizational mission, and one that captures the interests of individual

donors. This paper develops and examines the empirical usefulness of these competing models in the context of the OECD DAC peer review process.

The paper is organized as follows: In the subsequent section of the paper, we present background information on the OECD DAC and the peer review process. We then develop expectations about the competing hypotheses and specify measures that help us to distinguish between bureaucratic process factors and individual donor interests. The first set of empirical analyses focus on explanatory factors related to the assignment of peer examiners from 1962 to 2020. We find that the bureaucratic process model receives more empirical support than the individual donor interest model. The second set of empirical analyses examines the selection of recipient countries for in-depth field visits when the review process is ongoing. Due to OECD reporting standards, the temporal domain of the analyses ranges from 1996 to 2020. Here we find that, while still largely within the bureaucracy-driven guidelines for the peer review process, the donor interest model is relevant for this assignment decision. We conclude with a discussion of our findings and point to the next research steps.

2 The Donor Peer Review System in the OECD DAC

The Organisation for Economic Co-operation and Development’s (OECD) Development Assistance Committee (DAC) is a forum in which member states come together to “promote coordinated, innovative international action to accelerate progress towards the Sustainable Development Goals (SDGs) in developing countries and improve their financing.” Originally established in 1960 as the Development Assistance Group (DAG) within the OECD forerunner, the Organisation for European Economic Co-operation (OEEC), the Committee consisted of founding members Belgium, Canada, France, Germany, Italy, Portugal, the United Kingdom, the United States and the Commission of the European Economic Community; Japan was also invited to join, and the Netherlands became a member within the year. By 2021, DAC membership has grown to include a total of 29 countries,¹ in addition to the European Union—officially classified as a “partner,” but acting as a full member in the committee; as well as seven “participant” countries;² and six “observer” organizations.³ The OECD DAC prides itself on operating on the basis of consensus among all members but it also produces soft law instruments like the peer review that serves to shape the behavior of states. It thus becomes essential to understand better how the institution goes about synthesizing consensus around policies and how it monitors state compliance with such policies.

¹Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, The Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, United Kingdom, and United States.

²Azerbaijan, Bulgaria, Kuwait, Qatar, Romania, Saudi Arabia, and United Arab Emirates.

³Asian Development Bank, African Development Bank, Inter-American Development Bank, International Monetary Fund, United Nations Development Programme, and World Bank.

Consensus-building in the OECD DAC takes place as representatives from member states and staff from the Secretariat work together within both formal and ad hoc working groups. In the case of the DAC, the objective is to come to an agreement about best practices for development cooperation and foreign aid giving. The outcome of these processes is enshrined in more or less formalized decision products, ranging from official DAC recommendations and declarations to more informal documents detailing principles and guidelines on specific topics. Then, whatever the consensus understanding of best practices for development assistance is at a given point in time, this gets reflected in the monitoring and evaluation of DAC donors through the peer review mechanism.

Since the early 1960s, peer reviews of DAC members have taken place in regular intervals. At first, peer reviews were annual but became less frequent in the late 1970s. Today reviews of DAC members take place every four to six years. Each peer review consists of a team of two other DAC member states (referred to as “examiners” throughout the paper) and the OECD Secretariat that examine aid giving practices of the DAC member that is under review (referred to as the “reviewee” throughout the paper). The peer review process typically lasts several months, beginning with the donor country producing a Memorandum of Understanding (MoU) where it presents the state of its development cooperation efforts to the OECD Secretariat and the organization’s staff in charge of the review process. The peer review team not only travels to the reviewee’s capital to visit relevant development actors. It also travels to recipient countries to closely examine the reviewee’s aid activities and practices in the field—or, it examines field operations through meetings or calls with relevant stakeholders. The information and assessments conducted in the field are an integral part of the final peer review.

Once the peer review team has collected all the relevant information and has concluded the examination process, the Secretariat is in charge of drafting a document that conveys the findings of the review. This process may involve examiner feedback. Once drafted the report is then presented and discussed in a formal DAC meeting. Importantly, the report includes a section titled “DAC’s Main Findings and Recommendations,” or “Summary and Conclusions” in older documents, where an explicit set of recommendations are outlined for how the reviewed donors should proceed to improve its aid giving practices. These recommendations used to be fewer in number and less detailed in terms of precise metrics and expectations, but have grown more specific and thorough over time. During the formal peer review meeting, the reviewed donor has the chance to question and possibly even request changes to the report document before a final draft gets distributed to all member states and posted online.

In this paper, we focus on two key aspects of the peer review mechanism—i.e., the choice of examiner countries and the choice of the recipient(s) to feature in an in-depth assessment—as a way to study whether the peer review process is driven by political incentives or whether it is largely shaped by bureaucratic forces.

3 Business as Usual *vs* Donor Interests: Initial Expectations

Given that an OECD DAC member’s foreign aid program is to be evaluated, what determines which fellow member states will serve as peer examiners? And, how are one or more of the reviewee’s field operations chosen for field visits and special examination? These two choices may be politically motivated. For example, donors with certain ideological approaches to development may prefer to be examined by peers who share that approach. Or, donors may prefer to highlight recipient countries in which their aid programs work well. Alternatively, they may be the result of practical bureaucratic processes or standard-operating procedures that maximize efficiency criteria, including e.g. shared language for the choice of examiner, or that maximize the credibility of the peer review process, including e.g. the request to not conduct repeated field visit examinations in the same recipient country.

Our understanding of the OECD DAC peer review system is that there is a bureaucratic process, or institutional guidelines, that shape examiner choice—or the selection of which member states get to review whom. These guidelines seek to facilitate the practical business of conducting reviews and maintain its credibility. For instance, in order to serve as a examiner for a review, a member state needs to be available (e.g. not already be busy with another recent review). The peer examiner should be able to conduct the review efficiently (e.g., newer DAC members might serve as examiners alongside more experienced members). And, there should not be an obvious conflict of interest (e.g., EU member states may not serve as examiners for a peer review of EU’s aid giving). Yet, it may be plausible that DAC members might lobby the OECD Secretariat for examiner-reviewee pairings that will be beneficial for them. For example, we can imagine that a reviewee might have a preference for peer examiners that are ideologically or politically aligned with them. Such a politically motivated examiner-reviewee pairing might produce biased outcomes or recommendations of the peer review process insofar as the examiners might have incentives to portray a reviewee’s aid practices in a better light, or produce recommendations that are easy to implement or that might overlook malpractice.

Further, it is our understanding that the choice about which field operations to examine in-depth during the review process is also shaped by a bureaucratic process that maximizes the ease by which the review process is implemented as well as the credibility of the review process. For example, reviewees are asked not to select the same recipient countries across multiple reviews. Yet, here too, exists the possibility that reviewees lobby the secretariat or the peer review team directly, advocating for particular recipient countries with whom they have long-standing or more favorable relations and which might serve better to document good practices in aid giving. In our empirical analyses of these two assignment choices, we interpret statistically significant associations between politically salient factors and assignment choices as evidence that reviewees may indeed lobby the Secretariat for favorable assignments. On the contrary, if

politically salient factors are not systematic predictors of either choice, we would interpret this pattern as indicative that the review process is bureaucracy-driven.

Our expectations about the extent to which bureaucratic guidelines or politics shape assignments vary by the kind of assignment choice. We expect that the early stages of the process are likely to be driven by the Secretariat and the OECD team in charge of peer reviews. Thus, practical considerations about how to make the implementation of the review run smoothly should prevail over any attempt for donors to lobby for any particular examiner. Moreover, DAC donors are already a fairly homogeneous group of countries, so it is unlikely that any of them would go to great lengths to demand a particular choice of examiners. It follows that we expect practical considerations to dictate the process, including e.g. member state availability to review, balance in examiner pairs of experienced and newer donors, shared language. We expect political considerations like political alignment, trade relations, or alliances to be less consequential in the choice of peer examiners.

Once the review team is in place and the review process starts, there might be more room for politics to shape decisions. It is plausible, for example, that the reviewee has strong incentives to showcase its development cooperation program at its best—that is, in those recipient countries it works well with. Here, we expect donors to lobby the peer review team to visit recipients who are especially important to the donor, such as those receiving a large proportion of the donor’s aid budget, or those who used to be former colonies of the donor.

If factors related to the relationship between donor and recipient are strong predictors of recipient choice empirically, we would interpret this as evidence that the kinds of donor-recipient politics we know exist in the aid allocation process are likely to also exist at this stage of the peer review process. On the other hand, if factors related to the ease of working with a recipient from the perspective of the OECD as a whole—as opposed to the particular donor being reviewed—are strong predictors of the choice of recipients for in-depth evaluations, then we would take this as an indication that the peer review process remains Secretariat-driven throughout, and that there is little room for donors to (successfully) influence these key aspects of reviews.

To address these questions empirically, we first analyze the determinants of peer examiners for the full universe of peer reviews—516 reviews, 1962–2020. Second, we assess the choice of recipients who receive an in-depth assessment for the 116 recent reviews for which we have field visit information, 1996–2020.⁴

⁴See the Appendix for descriptive information about the peer review data.

4 Determinants of Peer Examiner Choice

In our empirical investigation of the role of OECD bureaucracy interests versus donor interests in the peer review process, we first turn to the decision of which member states serve as examiners in a given review.

4.1 Data and Methods

We construct a panel data set covering all 516 DAC peer reviews from 1962 to 2020 in which the unit of analysis is the pair of reviewed donor (“reviewee”) and peer examiner country (“examiner”) nested within each review. This approach ensures that we account for the potential interdependence of examiner choices in any given DAC review. We construct this data set in three steps. First, we identify the potential set of DAC peer examiners. A peer examiner must be a DAC member by the time the review is undertaken. Second, we compute all unique pairs of potential examiners. The set of potential examiners varies over time as the DAC has admitted new members throughout history. In our sample, the number of potential examiners ranges from 10 donors in 1962 up to 28 donors since 2016. Our panel data set is therefore unbalanced in the number of observations for any given review. Third, we add all relevant covariates—respectively for the reviewed donor, each peer examiner, the reviewee-examiner dyad, and the examiner dyad.

4.2 Dependent Variable

Our dependent variable is binary, capturing whether a given examiner pair has undertaken the review for a given reviewed donor and a given year. While the binary dependent variable would imply estimation of non-linear models, we rely on linear probability models. This affords us the possibility to include fixed effects at the review level, essentially controlling for any idiosyncratic effects of the review(ee). We are therefore left with examiner effects, reviewee-examiner effects, and examiner-pair effects to explain the choice of peer examiners.

4.3 Independent Variables

Our key predictors are organized into two sets and follow our main theoretical considerations for the choice of peer examiners. The first model is the “bureaucratic process” model, which captures the official criteria for examiner selection as developed in OECD DAC guidance. Although this guidance has continuously been updated over the years, long-standing criteria include that two DAC members act as peers for each review. These examiners are designated by the OECD Secretariat and are matched, where possible, with “reviews covering a member of similar size and complexity, and language and geography” (OECD 2021). Furthermore, examiners should not serve as an examiner of the same donor twice in a row.

We operationalize these criteria at the reviewee-examiner level as follows. To account for the criteria related to “size” and “complexity” we construct an indicator of being a major donor—France, Germany, Japan, United Kingdom, and United States—and include two indicators for whether (i) the review team includes at least one peer examiner in the same category as the reviewee, and (ii) both examiners are of the same type. We include a similar set of relational indicators for membership in the European Union (EU). These EU indicators are time-varying across reviews because countries joined the EU at different points in time. As further proxy for complexity, we add two relational indicators for whether at least one or both peer examiners have the same organizational model as the reviewed donor (DAC 2009). In addition, we include indicators for common language. That is, we measure whether at least one or both examiners speak the same language as the reviewed donor. Similarly, we capture whether at least one or both examiners are contiguous to the reviewed country (Mayer and Zignago 2011). Finally, we include an indicator for whether the same examiner pair was allocated to the reviewee at its previous review; as well as three indicators capturing whether at least one examiner or both—individually and as a pair—have reviewed the same reviewee before.

The second model is what we consider a “donor interest” model, which allows for politics to play a systematic role in the formation of review teams. In other words, states may hold particular preferences for being assigned certain peer examiners. We proxy donor interests through relational variables that have either been commonly used as proxies for donor interest in existing aid allocation models or that we develop specifically in the context of the DAC peer reviews. Our first relational variable measures foreign policy alignment between the reviewed donor and its peer examiners. Consistent with numerous previous studies, we proxy foreign policy alignment as the ideal-point distance of any of the two examiners and reviewee pairs based on their voting patterns in the UN General Assembly (Bailey, Strezhnev, and Voeten 2017). If OECD DAC members have a preference for reviewing countries that are closely (or not closely) aligned with their foreign policy and are, as a result, directly lobbying the OECD Secretariat to be paired with them, we would expect this measure to be a statistically significant predictor of a reviewee-examiner pair, although we remain agnostic about the direction of the statistical association. As we have two examiners assigned per review, we construct both the average and the range of the pairwise relational variable.

Our second variable is a proxy for power, measured by the Composite Index of National Capability (CINC) and including military expenses, military personnel, total population, urban population, energy consumption, and steel production (Singer 1972). In lobbying for being reviewed by a more powerful donor, a reviewee may increase the salience of the peer review. Increased salience may work in favor of reform-minded donor administrations which need the peer review verdict to overcome domestic opposition to reform (Porter and Webb 2007).

Turning to domestic politics, we posit that donors may prefer to be paired with examiners that have sim-

ilar domestic institutions and partisan orientations. It could be, for instance, that examiners and reviewees prefer to engage in peer review with more like-minded countries, or “sympathetic interlocutors,” in order to reduce the risk of unpredictable reviews. If DAC members have a systematic preference for reviewing and being reviewed by peers that are institutionally and ideologically similar, then they would directly lobby the OECD Secretariat to be paired with them. We consider two sets of domestic politics alignment variables. The first is the average difference of partisan ideology between reviewee-examiner pairs, as well as the range of these differences (Fuchs and Richert 2018). The second set contains the average and range over the differences in the index of deliberative democracy between reviewee-examiner pairs. The deliberative democracy index is a pertinent measure of the quality of democracy, which serves to address the limited variation of conventional democracy measures such as the polity index in the sample of advanced donor countries (Teorell et al. 2020).⁵

4.4 Control Variables

For our main analysis, we opt for parsimonious models that only include review fixed effects. This serves to capture any idiosyncratic effects of the review, the reviewee, and any year effects. In robustness tests, we remove fixed effects and instead estimate random-effects models, which assume a normally-distributed systematic error component across reviews and which allow us to include additional features of the review, such as the involvement of a third examiner and the number of available donors in the peer examiner pool. We cluster standard errors on reviews, considering the interdependent choices of examiner pairs within reviews. Unless otherwise stated, our sample excludes reviews in which the EU institutions are being reviewed, given that many covariates are undefined.

4.5 Main Results

Table 1 juxtaposes the two models that include variables that help explain the process of examiner assignment—the choice of examiner pairs—and whether this process is driven by factors related to an impartial bureaucracy or whether it follows donor interest. We proceed by first introducing each model separately, and then estimating a joint model.

In Table 1, Model 1, we probe the “bureaucratic process” model. We find that DAC peer review pairs are chosen such that at least one reviewee-examiner pair is comparable in terms of donor size ($p < 0.01$), while it tends to be unlikely that both pairs fulfill this criterion ($p < 0.1$). Substantively, if at least one pair has similar size, it is 0.54% more likely to be chosen for review compared to where both examiners are of different size

⁵Table A1 in the Appendix presents descriptive statistics of all variables in the analysis.

than the reviewee. We do not find significant effects for common membership in the EU, nor with respect to the independence of the aid agency. Finally, it is relatively more likely that both examiners are familiar with the reviewee due to being paired up previously with that reviewee ($p < 0.1$). Common language and geographic proximity are not statistically significant.

Table 1: Determinants of Peer Examiner Choice

	(1)		(2)		(3)	
At least one similar size	0.549***	(0.111)			0.625***	(0.139)
Both similar size	-0.114*	(0.067)			-0.111	(0.105)
At least one similar EU	0.065	(0.069)			0.092	(0.091)
Both similar EU	-0.070	(0.067)			-0.137*	(0.083)
At least one common language	0.089	(0.082)			0.192*	(0.105)
Both common language	0.252	(0.199)			0.313	(0.251)
At least one contiguous	-0.057	(0.084)			-0.209**	(0.104)
Both contiguous	-0.155	(0.181)			-0.090	(0.268)
At least one same model	-0.020	(0.056)			-0.072	(0.075)
Both same model	0.025	(0.095)			0.083	(0.126)
Review twice in a row	-0.085	(0.075)			-0.190**	(0.097)
Both before as pair	0.145	(0.134)			0.108	(0.164)
At least one before	0.001	(0.084)			0.006	(0.117)
Both before	0.125*	(0.065)			0.128	(0.088)
Average UNGA distance			-0.157	(0.134)	-0.026	(0.150)
Dispersion of UNGA distances			-0.102	(0.101)	-0.177*	(0.105)
Average power difference			0.696	(2.141)	-0.650	(2.151)
Dispersion in power difference			1.025	(1.516)	0.961	(1.541)
Average partisan distance			0.106	(0.140)	0.109	(0.138)
Dispersion of partisan distances			0.012	(0.116)	0.013	(0.116)
Average difference in democracy			0.042	(0.575)	-0.075	(0.583)
Dispersion democracy differences			-0.925**	(0.388)	-0.651	(0.395)
Observations	83646		52387		52387	
Within R2	0.001		0.000		0.001	

Linear regressions with review fixed effects and standard errors clustered on reviews in parentheses.
Significance levels: * $< .1$ ** $< .05$ *** $< .01$

Model 2 tests the “donor interest” model. The only statistically significant coefficient is for the range of the differences in deliberative democracy ($p < 0.05$). This means that highly unequal pairings of donors with respect to deliberative democracy are less common. The remaining donor alignment variables, such as UN General Assembly vote alignment, relative power, and partisan ideology are not systematically related to the choice of examiner pairs.

Model 3 pits the two sets of explanations against each other. This enhances the precision in the estimates for some covariates, while others result as not statistically significant. The robust predictors include the positive coefficient on at least one donor pair of similar size ($p < 0.01$), a weakly negative coefficient for both examiners being similar with respect to EU membership ($p < 0.1$), a weakly positive association for at least one examiner speaking the same language as the reviewee and a weakly negative association for one pair of countries being contiguous ($p < 0.1$), both donors being similar to the reviewee on aid agency independence ($p < 0.05$), and a weakly lower likelihood of the same examiner pair chosen twice in a row for the same

reviewee ($p < 0.1$). The only variable from the donor interest model that reaches marginal significance is the diversity in ideal-point distances: the likelihood of two examiners with opposing foreign policy preferences being chosen tends to be smaller ($p < 0.1$).

While our main models exclude the EU institutions as reviewee, we can use this special case to probe the extent to which examiner pair choice follows a bureaucratic process. Specifically, DAC rules specify that no EU member country reviews the EU institutions. This limits the examiner pool considerably, which is why we expect our bureaucratic process variables to matter less systematically than for ordinary DAC member states. In further analysis, we indeed verify that the EU institutions were not reviewed by any EU member states (with the exception of two exceptional cases among 516 reviews). Table 2 corroborates this result when controlling for other determinants of examiner choice. A review of the EU institutions is also less likely to involve two examiners who reviewed it before.

Table 2: Determinants of Examiner Peer Choice for Reviews of the European Union

	(4)		(5)	
	EU is reviewed		Other reviewees	
At least one major donor	0.079	(0.266)	0.183***	(0.065)
Both major donors	-0.397*	(0.223)	-0.458***	(0.144)
At least one EU member	-2.596***	(0.214)	0.112	(0.070)
Both EU members	-0.216*	(0.125)	0.069	(0.067)
At least one same model	-0.327	(0.250)	-0.011	(0.055)
Both same model	-0.023	(0.379)	0.029	(0.096)
Review twice in a row	-0.112	(0.385)	-0.056	(0.076)
Both before as pair	-0.188	(0.668)	0.166	(0.134)
At least one before	0.195	(0.311)	0.009	(0.084)
Both before	-0.671*	(0.370)	0.129*	(0.065)
Observations	4295		83646	
Within R2	0.016		0.000	

Linear regressions with review fixed effects. Standard errors clustered on reviews in parentheses. Significance levels: * $< .1$ ** $< .05$ *** $< .01$

In summary, we have found that the choice of peer examiners is broadly in line with a bureaucratic process model whereby examiner teams tend to include at least one donor of similar size and two donors with similar aid governance; examiners are chosen to avoid obvious conflicts of interest (such as EU institutions being reviewed by EU member states); and examiners who are familiar with the reviewee through prior review are more likely to be chosen again. In contrast, the empirical patterns are inconsistent with donor interest arguments as we find no evidence that examiner choice is driven by similar foreign policy preferences, political institutions, or partisan alignment.

We now probe some additional implications that should be observed if the choice of peer examiners is indeed driven by bureaucratic processes. Aside from practical considerations and concerns about the

credibility of the review process, one might also expect OECD DAC staff to be interested in ensuring that the peer review process is effective in getting states to adopt best practices. If this is true, staff might look for opportunities to make success more likely by promoting particular reviewee examiner pairings that increase the effectiveness of the review process. To this end, the DAC Secretariat could be sure to pick examiners which do well on widely-accepted “good aid” practices when putting together the peer review team for a reviewee that is susceptible to peer review pressure or where a change in aid policy would have great impact. We operationalize these conditions as follows: reviewed donors are likely are more amenable to peer pressure if they have an independent aid agency which impersonates the development interests within the government (Fuchs and Richert 2018). The extent to which reviews can achieve impact is captured by the major donor indicator. As indicators of good aid practice on the examiner side, we average across both examiners the total size of the aid budget and the ODA/GNI quota.

Table 3: Determinants of Peer Examiner Choice Depending on Reviewee and Examiner Characteristics

Examiner Characteristics	Reviewee Characteristics							
	Independent Agency		Non-Independent Agency		Major Donor		Non-Major Donor	
	(6)		(7)		(8)		(9)	
(Logged) average ODA	0.217**	(0.096)	-0.090	(0.086)	0.327**	(0.145)	-0.043	(0.075)
Average ODA/GNI (%)	0.809**	(0.366)	0.066	(0.309)	0.815	(0.510)	0.128	(0.257)
At least one similar size	0.458**	(0.206)	0.762***	(0.190)	0.713***	(0.254)	0.317	(0.213)
Both similar size	-0.094	(0.171)	-0.090	(0.144)	-1.087**	(0.477)	-0.105	(0.128)
At least one similar EU	0.041	(0.142)	0.146	(0.125)	0.180	(0.197)	0.043	(0.103)
Both similar EU	-0.052	(0.111)	-0.187	(0.121)	-0.260	(0.174)	-0.104	(0.094)
At least one common language	0.264*	(0.134)	0.197	(0.167)	0.355*	(0.180)	0.144	(0.128)
Both common language	0.361	(0.307)	0.398	(0.428)	0.025	(0.435)	0.459	(0.307)
At least one contiguous	-0.267*	(0.151)	-0.189	(0.145)	-0.050	(0.257)	-0.281**	(0.116)
Both contiguous	-0.291	(0.316)	-0.000	(0.417)	-0.548	(0.389)	0.096	(0.375)
At least one same model	-0.094	(0.107)	-0.105	(0.107)	-0.277*	(0.165)	-0.022	(0.083)
Both same model	0.022	(0.177)	0.136	(0.178)	0.178	(0.250)	0.050	(0.147)
Review twice in a row	-0.121	(0.146)	-0.218*	(0.131)	-0.331	(0.218)	-0.132	(0.108)
Both before as pair	0.069	(0.251)	0.084	(0.218)	-0.141	(0.370)	0.118	(0.181)
At least one before	0.019	(0.148)	-0.115	(0.187)	-0.148	(0.216)	-0.033	(0.144)
Both before	0.155	(0.142)	-0.037	(0.123)	0.299	(0.196)	-0.005	(0.105)
Average UNGA distance	0.378	(0.237)	-0.229	(0.207)	-0.018	(0.393)	0.022	(0.168)
Dispersion of UNGA distances	-0.093	(0.176)	-0.184	(0.143)	-0.366	(0.297)	-0.091	(0.115)
Average power difference	-10.159	(6.559)	3.017	(2.137)	7.694**	(3.463)	-2.028	(3.008)
Dispersion in power difference	-3.775	(3.346)	2.913*	(1.613)	4.328	(3.629)	0.045	(1.742)
Average partisan distance	0.078	(0.201)	0.091	(0.188)	0.088	(0.263)	0.119	(0.163)
Dispersion of partisan distances	-0.157	(0.163)	0.166	(0.165)	-0.022	(0.262)	0.030	(0.130)
Average difference in democracy	0.372	(0.846)	0.086	(0.815)	-0.381	(1.074)	0.287	(0.736)
Dispersion democracy differences	-1.218**	(0.508)	-0.515	(0.549)	-1.232**	(0.572)	-0.537	(0.504)
Observations	31577		44331		18000		63439	
Within-R2	0.002		0.000		0.004		0.000	

Linear regressions with review fixed effects. Standard errors clustered on reviews in parentheses.

Significance levels: * <.1 ** <.05 *** <.01

Table 3 corroborates our expectations about how the DAC Secretariat can make strategic use of peer examiners where reviews are under the condition to be more effective. Specifically, we find that reviewees with independent agencies are more likely to be paired with examiners with large aid budgets, compared to

reviewees without such agencies. We also find that major donors are reviewed by examiners with large aid budgets and relatively higher ODA/GNI quotas. These findings are consistent with the strategic anticipation of an effective peer review.

On the other hand, if peer review choices are driven by bureaucratic process, we should not observe donor interest variables to matter even when and where donor interests are most salient. In other words, to further refute donor interest explanations, we should also dismiss the possibility that donor interest may be a contextually important determinant of peer examiner choice. To that end, we consider two kinds of situations in which donor interest variables should be particularly salient. One is when the reviewee occupies a global position of influence, notably as temporary member of the UN Security Council. Because this role puts them in the spotlight, governments will be sensitive to critical peer reviews. They should therefore have an interest to get examiners that are close to their own foreign policy preferences and that face similar domestic policymaking challenges. Second, in an election year, government policy is under increased scrutiny from voters. Therefore, reviewees have a preference for examiners with similar preferences and domestic policymaking challenges.

Table 4: Determinants of Peer Examiner Choice Depending on Reviewee and Examiner Characteristics

Examiner Characteristics	Reviewee Characteristics							
	UNSC member (10)		Non-UNSC member (11)		Election (12)		No Election (13)	
(Logged) average ODA	-0.303	(0.223)	0.021	(0.076)	0.087	(0.119)	0.033	(0.073)
Average ODA/GNI (%)	0.635	(0.725)	-0.004	(0.264)	-0.022	(0.395)	0.413	(0.282)
At least one similar size	0.847	(0.583)	0.381*	(0.206)	0.703***	(0.226)	0.557***	(0.179)
Both similar size	-0.413	(0.333)	-0.021	(0.132)	-0.204	(0.195)	0.001	(0.135)
At least one similar EU	0.464*	(0.237)	-0.038	(0.113)	0.180	(0.160)	0.055	(0.112)
Both similar EU	-0.267	(0.272)	-0.086	(0.097)	-0.337**	(0.135)	-0.007	(0.104)
At least one common language	0.232	(0.456)	0.134	(0.129)	0.079	(0.182)	0.284**	(0.129)
Both common language	-0.040	(0.800)	0.471	(0.321)	-0.072	(0.364)	0.510	(0.330)
At least one contiguous	-0.817**	(0.312)	-0.188	(0.119)	-0.210	(0.202)	-0.226*	(0.121)
Both contiguous	-0.211	(0.252)	0.146	(0.380)	-0.356	(0.385)	0.008	(0.345)
At least one same model	-0.159	(0.237)	-0.007	(0.086)	-0.088	(0.125)	-0.076	(0.093)
Both same model	0.128	(0.414)	0.034	(0.150)	0.336	(0.235)	-0.077	(0.145)
Review twice in a row	-0.618**	(0.283)	-0.093	(0.112)	-0.302*	(0.161)	-0.108	(0.123)
Both before as pair	0.626	(0.619)	0.044	(0.183)	-0.287	(0.210)	0.322	(0.225)
At least one before	0.092	(0.503)	-0.036	(0.143)	0.067	(0.217)	-0.062	(0.143)
Both before	-0.029	(0.265)	0.054	(0.111)	-0.145	(0.157)	0.234**	(0.112)
Average UNGA distance	0.242	(0.370)	-0.076	(0.192)	-0.072	(0.245)	0.032	(0.198)
Dispersion of UNGA distances	-0.253	(0.301)	-0.080	(0.123)	-0.494***	(0.167)	0.060	(0.140)
Average power difference	-15.221	(9.172)	0.397	(3.064)	-1.766	(4.721)	0.739	(2.878)
Dispersion in power difference	4.996	(5.862)	-0.498	(1.751)	2.272	(2.888)	-0.041	(1.767)
Average partisan distance	0.126	(0.499)	0.126	(0.163)	-0.027	(0.241)	0.166	(0.168)
Dispersion of partisan distances	0.347	(0.400)	-0.019	(0.132)	0.054	(0.201)	-0.009	(0.143)
Average difference in democracy	-1.959	(1.208)	0.619	(0.837)	-0.586	(0.819)	0.461	(0.865)
Dispersion democracy differences	1.643	(1.194)	-0.894	(0.542)	-0.172	(0.592)	-1.067*	(0.548)
Observations	5617		35638		18435		33952	
Within-R2	0.005		0.001		0.002		0.002	

Linear regressions with review fixed effects. Standard errors clustered on reviews in parentheses.

Significance levels: * <.1 ** <.05 *** <.01

Table 4 shows the results. We do not find any significant differences in donor interest variables during times of increased donor sensitivity to critical reviews compared to ordinary times. We take these results as further suggestive evidence that donor politics does not matter for the assignment of peer examiners.

4.6 Robustness Tests

In the supplementary appendix, we establish the robustness of our findings in various ways. We begin by expanding the set of variables proxying for donor politics to include arms trade, military alliances, and diplomatic representation (Terman and Voeten 2018). We measure whether at least one pair or both pairs trade arms with each other (SIPRI 2021), whether they are members of the same alliance (Leeds et al. 2002), or whether they have mutual diplomatic representation (Bayer 2006), respectively. Neither are any of these variables robustly significant nor do they affect our main results (Table A2).

Next, we construct a time-varying measure of major donorship by considering the top-5 donors in every year in terms of their gross domestic product. The resulting relational indicators are highly correlated with our related static measure. Using this alternative operationalization, we find that examiner teams are significantly more likely to include at least one reviewee-examiner pair of similar size (Table A3).

We also use a different time-varying measure for relative donor size based on aid budgets. Specifically, we compute the ratios of the logged ODA budgets in each reviewee-examiner pair and include the geometric mean of these ratios. We also calculate the dispersion of ODA budget differences, by dividing the larger ratio by the smaller ratio. Consistent with our earlier findings based on major donors, we find that a scenario in which the reviewee is a large donor but both examiners are small donors is relatively unlikely ($p < 0.01$), and that there tends to be a weakly positive effect of relational size dispersion (Table A4).

We address possible concerns with the model specification in our main analyses by dropping review fixed effects and instead estimating random-effects regressions. This assumes a latent factor that is normally distributed across reviews, which is not implausible given the high number of clusters. The results are qualitatively similar. Reviews are more likely include one major power (but not two major powers), and more likely draw on examiners who are both familiar with the reviewee. Furthermore, we continue to find a positive effect of average donor size, and a negative effect of dispersion in differences of deliberative democracy (Table A5).

Again addressing possible specification concerns, we re-run our main analyses clustering standard errors on years. This is to address the potential interdependence of peer reviews in the same year. For example, it is unlikely that a given examiner pair will undertake two reviews in the same year, given that the reviews commit administrative resources in the donor. When clustering standard errors this way, our results are

virtually unaffected (Table A6).

We probe additional examiner characteristics—unrelated to the reviewed donor—as drivers of examiner choice. While these characteristics are not absorbed by fixed effects, they also provide another avenue into testing for donor influence effects with respect to the examiners. For example, we might expect that examiners are not called upon if they face dire economic circumstances, such as high unemployment and low growth, and if they have lower per capita income. In addition, we consider that certain political characteristics affect the likelihood of being chosen for examiner, specifically the extent of corruption. We retain the most unfavorable value of each covariate from both examiners, expecting a negative sign with the likelihood of being chosen as examiner. We also include the average gap in total ODA/GNI across both pairs. None of the variables result as consistently statistically significant (Table A7).

Finally, in further analysis, we probe additional potential determinants that do not neatly fit into either bureaucratic process or donor politics but have precedent in the literature. Specifically, we probe whether alignment in donor political economies affects examiner selection (Dietrich 2021). For example, it may facilitate the review process and increase review quality if examiners can assume domestic politics to be working similarly in the reviewee as in their own countries. From a donor interest perspective, countries of similar political economy may be more ideologically aligned and therefore less likely to question certain aid practices. We also test whether reviewing schedules are driven by reciprocity, such that a given pair of donors reviews each other in direct succession. Such role-switching could be evidence of bureaucratic efficiency given that the same individuals would be involved in the review discussions. However, reciprocal reviews could also be in line with donor interest given that donors may be friendlier to each other considering potential for immediate retaliation if a first review was too critical. Empirically, we do not find alignment in donor political economies to matter. In contrast, direct reciprocity is a strongly significant predictor of peer review that operates independently of the ones discussed earlier (Table A8).

In conclusion, we have found robust evidence that the choice of peer examiners is more in line with an bureaucratic process model than a donor interest model. Across different model specifications, we found that examiner teams tend to have at least one major donor but are less likely to include two major donors, while including examiners that are both familiar with the reviewee from previous reviewing. In contrast, none of the donor interest variables is robust across model specifications.

5 Determinants of Recipient Country Visits

In the second part of this paper, we investigate the extent to which politics shapes the selection of recipients for in-depth assessments that are part of nearly every OECD DAC peer review and that make it into the final

review reports. The review process entails an assessment of the donor’s development assistance programming in the field, visiting between one or more recipient countries. Of the 121 peer reviews (1996–2020) for which we have access to the full report document, 116 (about 96%) include at least one recipient-specific assessment; 55 include at least two recipient-specific assessments; five have at least three; and one has information on four such assessments. This results in 177 recipient-specific assessments over 116 reviews.

Such recipient-specific assessments typically occur in the context of a short field visit in the recipient country and are undertaken by the peer review team.⁶ In fact, 162 of the 177 recipient-specific assessments in our sample (about 92%) takes the form of a field visit, which is subsequently summarized in a report that serves as an annex to the final peer review report. When a physical field visit is not possible, the peer review commission evaluates the donor’s field program through phone interviews with relevant actors (in four instances in our sample) or by inviting relevant actors from the field to answer questions about recipient country programs to examiner capitals (in seven instances), or by meetings and calls from other locations (three instances). In one case, a field visit to the recipient country was scheduled for after the completion date of the peer review report.

What determines the choice of recipient country for such in-depth assessments for the purpose of a peer review? According to official OECD peer review methodology,⁷ input from donors is explicitly included at this stage, as the reviewed member is asked to provide suggestions of possible recipients to consider for field visits to the review team. The guideline states that, “in the interests of transparency and objectivity,” members are asked to provide multiple options⁸ that satisfy a set of criteria—which the review team will also use in making the final selection:

“In selecting options for partner country visits, the reviewed member should consider the following criteria. (i) The country should be a significant partner for the reviewed member in terms of relationship, and level and scope of engagement; (ii) the country should be representative of the reviewed member’s programme for accountability purposes; (iii) it should provide an opportunity for the review team to reflect on the reviewed member’s application of its development co-operation policies and operational guidance; (iv) the programme should also provide for other learning opportunities (such as those related to global issues, exit strategies and aid management issues); and (v) the country should not have been visited by a peer review team in the same or the previous year.”

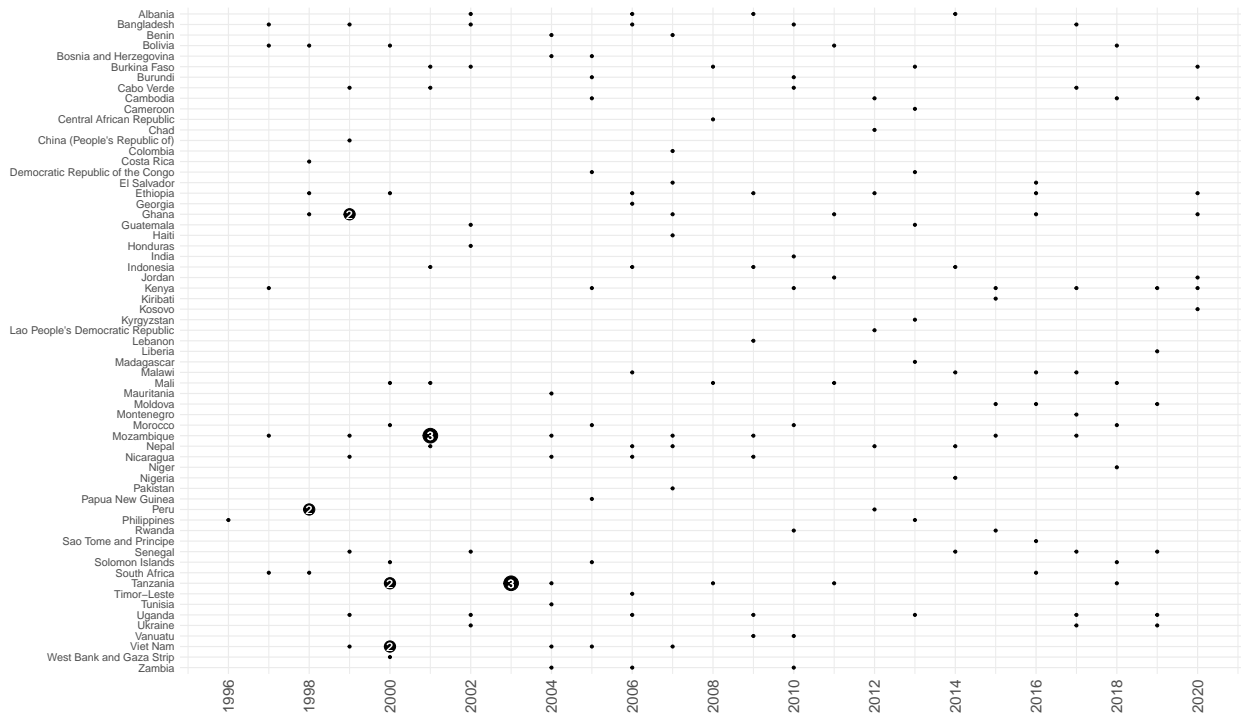
⁶Remember that a peer review team is composed of representatives from two other DAC donors (referred to as the two peer “examiners”), the DAC Secretariat, and possibly an observer; although it is unclear whether every member of the team participates in every recipient field visit or recipient assessment meeting.

⁷See for example document DCD-DAC(2010)19-FINAL dated March 20, 2017, and titled, “Guidance for Selecting Partner Countries to Visit as Part of the Peer Review Process.”

⁸“The review team should be offered a choice of three countries in cases where there is only one visit to a partner country and a choice of six countries in cases where the Peer Review includes two visits.”

Before addressing the first four criteria, which are less specific and leave some room for interpretation, we look at whether the fifth and last criterion is respected in the choice of recipient visits. Figure 1 lists all recipients who were the subject of at least one field visit in the sample of 116 peer reviews for which we have information about their recipient-specific assessments.⁹ Note that, contrary to what outlined in the peer review methodology, it is not uncommon for the same recipient to receive visits for reviews in back-to-back years—this happened as recently as Kenya being visited both for a 2019 review and for a 2020 review. In six instances, the same recipient was visited for two to three reviews in the same year, although this has not happened since 2003.

Figure 1: Recipient Visit Count by Year



With respect to the other, less specific criteria, while the reviewed donor suggests a list of possible recipients for field visits, it is the peer review team that ultimately makes the decision of where to go. Thus, on the one hand, we expect reviewed donors to want to showcase their programs in recipients that they already favor, with whom they have long-standing relationships, or which are strategically important.

On the other hand, we expect the peer review team to prefer field assessments to take place in countries that are easy to navigate and with infrastructure in place that facilitates evaluation. This could include

⁹See the Appendix for descriptive information about the peer review data.

countries that are relatively better developed, that score more highly on governance indicators, and that are more democratic. The OECD interest might also favor recipients who are particularly important not to the reviewee donor in particular, but to the group of DAC donors as a whole.

We test these expectations on the same sample of 116 peer reviews for which we have information about their recipient-specific assessments.

5.1 Data and Methods

We construct a review-donor-recipient data set covering 116 DAC peer reviews from 1996 to 2020. We identify the potential set of recipients that could be assessed for any review as any country who has appeared as a recipient of the reviewed donor’s ODA flows in any of the three years prior and up to the year of the peer review, according to the OECD DAC Creditor Reporting System (CRS) data. Each observation in the data is therefore a triad composed of (i) the donor which was subject to peer review, (ii) the year in which it was reviewed, and (iii) a potential recipient of that donor’s ODA which could have been chosen for a recipient-specific assessment.

5.2 Dependent Variable

We construct our dependent variable, the implementation of an in-depth recipient country assessment, by creating an indicator that assigns the value of one to recipients that were subject of a recipient evaluation for a donor-year review and zero to recipients who were not assessed through field visits. Given that certain reviews have more than one recipient-specific assessment, this indicator may be equal to one for one or up to four observations per review.¹⁰ The five reviews in our sample which have no recipient-specific assessments are not included in the analyses.

5.3 Independent Variables

To capture the idea of politics driving this selection, we include variables that measure the importance of the recipient to the reviewed donor (“reviewee”). First, we add a variable that measures the average amount of ODA that the reviewee spent on each potential candidate for field visits in the three years leading up to the year of the peer review. We also include a measure of the amount of ODA committed to candidate recipients for field visits as a percentage of the reviewee’s total aid budget. This measure aims to capture the importance of each recipient in the donor’s aid portfolio. Lastly, we include an indicator of colonial history, equal to one for recipients who were primarily a former colony of the donor and zero otherwise.

¹⁰As discussed, 116 reviews 1996–2020 include at least one recipient-specific assessment; 55 include at least two recipient-specific assessments; five have at least three; and one has information on four such assessments.

In terms of recipient characteristics, we include a measure of democracy (Polity V); government effectiveness, rule of law, and political stability and absence of violence/terrorism (all from the World Bank’s Governance Indicators); and life expectancy and GDP per capita (from the World Bank’s World Development Indicators). We also include the World Bank’s Statistical Capacity Indicator to capture the extent to which a given recipient can be expected to keep good records and thus facilitate operations and learning during a peer review; as well as a measure of international tourism (from the World Tourism Organization) to proxy for the ease of access of a given recipient. All of these recipient-country characteristics are included in our models for the year prior to the peer review.

To capture a recipient’s importance to the eyes of the OECD DAC as a whole—as opposed to the individual reviewee donor—we include the amount of aid committed to a recipient in a given year by all DAC donors together, as a percentage of the total amount committed by all DAC donors to all recipients in that year. Lastly, we include whether a recipient holds one of the rotating positions in the UN Security Council in the year of the review.¹¹

5.4 Control Variables

In some of our model specifications, we control for peer-review format, which allows us to account for the fact that donor peer reviews change formats over time; and for the size of the pool of potential recipients, given that it depends on how many recipients each donor has in its portfolio at a given time.

5.5 Main Results

Similar to the previous set of analyses, we estimate linear regression models with a binary outcome variable. Table 5 shows the results of the main specifications in our analyses. In the first column, we present results from a random-effects model without controls. The coefficient for the variable that proxies recipient importance to the donor—percent of donor ODA budgets given to the recipient—is positive and statistically significant. This suggests that recipients who are more important to donors are more likely to be targeted for in-depth program assessments. This idea is supported by the positive and significant coefficient for the indicator of colonial history, which suggests that being a reviewee’s former colony increases the probability that a recipient is selected for an in-depth assessment.

Notably, the coefficient for the variable that proxies recipient importance to the DAC as a whole—percent of cumulative ODA budget of all DAC donors given to the recipient—is also statistically significant, although smaller in magnitude; and the results also show evidence of a positive relationship with the statistical capacity

¹¹See the Appendix for a detailed description of all data sources.

of recipient countries. The former suggests that potential recipient’s relationship to the broader community of DAC donors also matters; while the latter can be interpreted as evidence of the peer review team favoring recipients who keep good records and make assessment and learning easier.

Another significant predictor in this first model is GDP per capita, with a negative coefficient. These results remain similar when controlling for peer-review format and recipient pool size, as presented in the second column of Table 5. In the third model presented in Table 5, we include recipient fixed effects. With this specification, we estimate coefficients from variation within each candidate recipient and thus account for recipient-specific characteristics that are not captured by the variables included in our model. This third specification also results in positive and significant coefficients for the variables that capture a recipient country’s importance to the reviewee—with effect sizes that are comparable to the previous two specifications. The coefficients for GDP per capita and statistical capacity are not significant in this specification.

Table 5: Determinants of Recipient Choice

	(14)	(15)	(16)
Aid from Donor, 3-yr Avg.	0.000013 (0.000017)	0.000013 (0.000018)	0.000015 (0.000018)
Aid from Donor, % of Donor Tot.	0.015571*** (0.001276)	0.015574*** (0.001295)	0.015270*** (0.001327)
Aid from all DAC Donors, % of DAC Donors Tot.	0.006890* (0.002707)	0.006941* (0.002711)	0.013022** (0.004081)
Former Colony of Donor	0.036195*** (0.009214)	0.035857*** (0.009251)	0.036959*** (0.009319)
Democracy, Prev. Year	0.000326 (0.000361)	0.000326 (0.000361)	0.000104 (0.001044)
Government Effectiveness, Prev. Year	0.004632 (0.007538)	0.004595 (0.007586)	0.006325 (0.013905)
Rule of Law, Prev. Year	0.011447 (0.007093)	0.011397 (0.007122)	0.003474 (0.016020)
Political Stability, Prev. Year	0.005273 (0.003098)	0.005244 (0.003107)	0.006663 (0.006218)
Life expectancy at birth, Prev. Year	0.000066 (0.000273)	0.000051 (0.000280)	0.000915 (0.001185)
GDP per capita, Prev. Year	0.000003*** (0.000001)	0.000003*** (0.000001)	0.000001 (0.000004)
UNSC Membership	0.007008 (0.006902)	0.007041 (0.006911)	0.005195 (0.007458)
International tourism, Prev. Year	0.000000 (0.000000)	0.000000 (0.000000)	0.000000 (0.000000)
Statistical capacity, Prev. Year	0.000366* (0.000168)	0.000362* (0.000169)	0.000006 (0.000356)
Effects	Random	Random	Recipient FEs
Controls	None	Doc Format and Pool Size	Doc Format and Pool Size
Num.Obs.	5,120	5,120	5,120
R2	0.050	0.050	0.041
R2 Adj.	0.048	0.047	0.020

Linear regressions, standard errors in parentheses. Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001.

While preliminary, these results suggest that politics matters for the stage in the peer review process where recipient countries are selected for an in-depth evaluation of the reviewed donor’s aid program. Recipients who matter relatively more to the reviewed donor are more likely to be chosen for these assessments. Yet, we know from the guidelines of the peer review process that donor input is explicitly welcomed at this stage.

The fact that recipients with higher GDP per capita and better statistical capacity appear to be favored, as well as recipients who matter relatively more to the community of DAC donors as a whole, would further suggest that peer review teams are guided by process-driven concerns in choosing recipients who arguably facilitate the implementation of an in-depth assessment. Yet other characteristics capturing the same concept—i.e., being democratic, stable, a member of the UNSC, or well-governed—do not appear to make recipients more likely to be selected for field visits.

It is also important to note that the magnitude of these effects is not very large. Substantively, recipient countries who are former colonies of the reviewed donors are about 3.6 percent more likely to be chosen for a visit than those who are not—all else equal; and a one-percent increase in the amount of the reviewed donor’s aid that the recipient gets makes such recipient less than 2 percent more likely to be chosen. The effect of a recipient’s relative importance to DAC donors altogether, of its GDP per capita, and of its statistical capacity are even smaller.

We interpret these findings to suggest that, while the bureaucratic process set in place by the OECD Secretariat drives this step of the peer reviews as well, reviewed donors are also successful in getting the peer review team to focus on recipients they care about most. The positive and significant results on statistical capacity and GDP per capita, however, also suggest that recipients who are able to facilitate learning (and potentially already performing well on long-standing development metrics important to the international community) are more likely to be chosen—regardless of what donor is being reviewed.¹²

6 Conclusion

In this paper, we systematically examined factors that shape important decisions in the implementation of the OECD DAC’s peer review mechanism. In a first instance, we looked at the process of assigning examiner countries, where donors are asked to assess aid practices of a DAC member that is to be reviewed. By assessing the empirical usefulness of two competing explanations for explaining this decision we found more evidence in support for a “bureaucratic process” model and less support for a model that captures donor interest. In a second instance we test the usefulness of the two models for explaining the selection of recipient

¹²The results presented and described in this section are largely the same if we limit our sample to reviews of traditional donors only—meaning those donors who joined the OECD DAC before 2000 and thus excluding South Korea, Iceland, and the East European donors who joined in the 2010s. More details on this in the Appendix.

countries to undergo in-depth assessments of donor practice. Our preliminary findings indicate that, while still largely within the guidelines for the peer review process, the donor interest model is relevant for this assignment decision.

The findings are preliminary insofar as we expect to revise and refine our models that capture bureaucratic or donor interest. To this end, we expect to conduct additional qualitative interviews with members of the OECD bureaucracy to be able to expand the set of factors that could proxy the bureaucratic process logic behind examiner assignment. Also, we also plan to expand the sample of reviews that contains information on recipient country visits. Up until now our sample was limited because the reporting format for the years outside of our sample was such that the information about recipient country examinations and field visits was not readily available. We expect to receive this information in the near future from the OECD Secretariat.

Finally, our preliminary findings have implications for scholars that explore compliance with peer review recommendations. Our results indicate that the assignment mechanism are not best explained by political considerations but rather by process-driven considerations. This suggests that questions that assess a country's compliance with recommendations should not be tainted by politically motivated selection effects.

References

1. Abbott, K., & Snidal, D. (1998). "Why States Act through Formal International Organizations." *Journal of Conflict Resolution* 42(1): 3–32.
2. Abbott, K., & Snidal, D. (2013). "Taking Responsive Regulation Transnational: Strategies for International Organizations." *Regulation and Governance* 7(1): 95–113.
3. Albaret, M. (2016). "Multilateralism under Transformation." *Routledge Handbook of International Organization* (10511). <http://www.routledgehandbooks.com/doi/10.4324/9780203405345.ch38>.
4. Alter, K., and Meunier. (2009). "The Politics of International Regime Complexity." *Perspectives on Politics* 7(1): 13–24.
5. Ashoff, G. (2013). 50 Years of Peer Reviews by the OECD's Development Assistance Committee: An Instrument of Quality Assurance and Mutual Learning. German Development Institute.
6. Bailey, M. A., Strezhnev, A., & Voeten, E. (2017). Estimating dynamic state preferences from United Nations voting data. *Journal of Conflict Resolution*, 61(2), 430-456.
7. Barnett, M. N., & Finnemore, M. (1999). The Politics, Power, and Pathologies of International Organizations. *International Organization*, 53(4), 699–732.
8. Bayer, R. (2006). Diplomatic Exchange Data set, v2006-1. <http://correlatesofwar.org> (Accessed 15 June 2021).
9. Broome, A., Homolar, A., & Kranke, M. (2018). Bad science: International organizations and the indirect power of global benchmarking. *European Journal of International Relations*, 24(3), 514–539.
10. Brown, R. (2010). "Measuring Delegation." *Review of International Organizations* 5(2): 141–75.
11. Carcelli, S. (2021). "Bureaucratic Structure and Treaty Compliance." Working Paper.
12. Carraro, V., Conzelmann, T., & Jongen, H. (2019). Fears of peers? Explaining peer and public shaming in global governance. *Cooperation and Conflict*.
13. Conzelmann, T. (2014). "The Politics of Peer Reviewing: Comparing the OECD and the EU." In *The Politics of Informantion. The Case of the European Union*, ed. T et al. Blom. , 49–62.
14. Dahlberg, Stefan, Aksel Sundström, Sören Holmberg, Bo Rothstein, Natalia Alvarado Pachon, and Cem Mert Dalli. 2021. The Quality of Government Basic Dataset, version Jan21. University of Gothenburg: The Quality of Government Institute.
15. Dietrich, S. (2021). *States, Markets, and Foreign Aid*. Cambridge: Cambridge University Press.
16. Dijkstra, Hylke. "Shadow bureaucracies and the unilateral control of international secretariats: Insights from UN peacekeeping." *The Review of International Organizations* 10, no. 1 (2015): 23-41.
17. Dingwerth, K., Schmidtke, H., & Weise, T. (2011). "Speaking Democracy: Why International Organizations Adopt a Democratic Rhetoric." In *Legitimacy and Legitimation in Global Governance*, eds. Jonas Tallberg and Michael Zürn.
18. Doshi, R., Kelley, J. G., & Simmons, B. A. (2019). The power of ranking: The ease of doing business indicator and global regulatory behavior. *International Organization*, 73(3), 611–643.
19. Dreher, Axel, Valentin Lang, B. Peter Rosendorff, James Raymond Vreeland, 2018, Buying Votes and International Organizations: The Dirty Work-Hypothesis, CEPR Discussion Paper No. 13290.
20. Dreher, Axel, Jan-Egbert Sturm, and James Vreeland, 2009, Development Aid and International Politics: Does membership on the UN Security Council influence World Bank decisions? *Journal of Development Economics* 88: 1-18.

21. Eyben, R. (2013). Struggles in Paris: The DAC and the purposes of development aid. *European Journal of Development Research*, 25(1), 78–91.
22. Farrell, H., & Newman, A. (2014). “Domestic Institutions beyond the Nation-State: Charting the New Interdependence Approach.” *World Politics* 66(2): 331–63.
23. Flüglister, K. (2012). Where Does Diffusion Take Place? The Role of Intergovernmental Cooperation for Policy Diffusion in Switzerland. *European Journal of International Relations*, 51(1), 316–349.
24. Freyburg, T., Krizic, I., Lavenex, S., & O’Flynn, C. (2017). *Democratic Governance and Transgovernmental Networks*.
25. Fuchs, A., & Richert, K. (2018). Development minister characteristics and aid giving. *European Journal of Political Economy*, 53, 186–204.
26. Groenendijk, N. (2009). EU and OECD benchmarking and peer review compared. Third EUCE Annual Conference: The EU in a Comparative Perspective, 1–18.
27. Hale, T., Held, D., & Young, K. (2013). “Gridlock: From Self-Reinforcing Interdependence to Second-Order Cooperation Problems.” *Global Policy* 4(3): 223–35.
28. Honig, D., & Weaver, C. (2019). “A Race to the Top? The Aid Transparency Index and the Social Power of Global Performance Indicators.” *International Organization*, 73(3), 579–610.
29. Hooghe, L., & Marks, G. (2015). “Delegation and Pooling in International Organizations.” *Review of International Organizations* 10(3): 305–28.
30. Johnson, Tana. “Organizational Progeny: Why Governments Are Losing Control Over the Proliferating Structures of Global Governance.” *Transformations in Governance*, 2014.
31. Kelley, J. G., & Simmons, B. A. (2015). “Politics by number: Indicators as social pressure in international relations.” *American Journal of Political Science*, 59(1), 55–70.
32. Keohane, R., & Nye, J. (1977). *Power and Interdependence: World Politics in Transition*. New York: Longman.
33. Leeds, B., Ritter, J., Mitchell, S., & Long, A. (2002). “Alliance Treaty Obligations and Provisions, 1815–1944.” *International Interactions*, 28(3), 237–260.
34. Lehtonen, M. (2007). “Environmental policy integration through OECD peer reviews: Integrating the economy with the environment or the environment with the economy?” *Environmental Politics*, 16(1), 15–35.
35. Liverani, A., & Lundgren, H. E. (2007). “Evaluation Systems in Development Aid Agencies: An Analysis of DAC Peer Reviews 1996–2004.” *Evaluation*, 13(2), 241–256.
36. Mahon, R., & McBride, S. (2009). “Standardizing and disseminating knowledge: The role of the OECD in global governance.” *European Political Science Review*, 1(1), 83–101. Mayer, T., & Zignago, S. (2011). “Notes on CEPII’s distances measures: The GeoDist database.”
37. McVoy, E. C. (1940). “Patterns of diffusion in the United States.” *American Sociological Review*, 5(2), 219–227.
38. Meyer, J. W., Drori, G. S., & Hwang, H. (2006). “World Society and the Proliferation of Formal Organization.” *Globalization and Organization: World Society and Organizational Change*, pp. 25–49.
39. Moravcsik, A. (2004). “Is There a ‘Democratic Deficit’ in World Politics? A Framework for Analysis.” *Government and Opposition* 39(2): 19–36.
40. Niemann, D., & Martens, K. (2018). “Soft governance by hard fact? The OECD as a knowledge broker in education policy.” *Global Social Policy*, 18(3), 267–283.

41. Nielson, D., & Tierney, M. (2003). "Delegation to International Organizations: Agency Theory and World Bank Environmental Reform." *International Organization* 57(2): 241–76.
42. Organisation for Economic Co-operation and Development's (OECD). 2021. *DAC Peer Review Methodology, Updated 2021*.
43. Pagani, F. (2002). "Peer Review As a Tool for Co-Operation and Change." *African Security Review* 11(4): 15–24.
44. Paulo, S., & Reisen, H. (2010). "Eastern Donors and Western Soft Law: Towards a DAC Donor Peer Review of China and India?" *Development Policy Review* 28(5): 535–52.
45. Porter, T., & Webb, M. (2007). The Role of the OECD in the Orchestration of Global Knowledge Networks. Canadian Political Science Association Annual Meetings.
46. Prakash, A., & Potoski, M. 2016. "Dysfunctional Institutions? Toward a New Agenda in Governance Studies." *Regulation and Governance* 10(2): 115–25.
47. Raustiala, K., & Victor, D. (2004). "The Regime Complex for Plant Genetic Resources." *International Organization* 58(2): 277–309.
48. Singer, J. D. (1972). "The "Correlates of War" project: Interim report and rationale." *World Politics*, 24(2), 243-270.
49. SIPRI (2021). SIPRI Arms Transfers Database (TIV of arms exports from all). <http://www.sipri.org/databases/armstransfers/sources-and-methods/> (accessed 15 June 2021).
50. Teorell, J., Kumlin, S., Sundström, A., Holmberg, S., Rothstein, B., Alvarado Pachon, N., & Dalli, C. M. (2020). The Quality of Government Dataset, version Jan20. University of Gothenburg: The Quality of Government Institute.
51. Vabulas, F., & Snidal, D. (2013). "Organization without Delegation: Informal Intergovernmental Organizations (IIGOs) and the Spectrum of Intergovernmental Arrangements." *Review of International Organizations* 8(2): 193–220.
52. Weaver, C. E. (2008). *The World Bank's Hypocrisy and the Poverty of Reform*. Ithaca: Cornell University Press.
53. Webb, M. C. (2004). Defining the boundaries of legitimate state practice: Norms, transnational actors and the OECD's project on harmful tax competition. *Review of International Political Economy*, 11(4), 787–827.
54. Westerwinter, O. (2019). *Transnational Public-Private Governance Initiatives in World Politics: Introducing a New Dataset*. St. Gallen.

Appendix

A1 Consensus Building Through OECD Working Groups

As discussed in the paper, when performing peer reviews, the OECD DAC monitors donor performance according to a set of (more or less explicitly) established ideas about what are to be considered best and worst practices in foreign aid giving. But how do these ideas become a shared consensus in the OECD DAC in the first place? From background research and interviews with organization officials, we understand the process of norm formation in the OECD as originating in one of three possible ways. First, one or more member states might care about an issue and take the initiative to start (and even fund) a working group for all members to get together and open a conversation on the topic. Alternatively, it is also possible that the OECD Secretariat might see an area where developing a norm would be useful, and bring it to the DAC to initiate discussions about it. Lastly, it is also possible that an event or scandal external to the OECD might gain the attention of the international development community and prompt the DAC and its members to react (e.g. Oxfam scandal).

From there, discussion around the potential norm is carried out in a working group or a subgroup. There are currently six main formal groups—the Network on Gender Equality (GENDERNET), the Network on Governance (GOVNET), the Network on Environment and Development Co-operation (ENVIRONET), the Working Party on Development Finance Statistics, the International Network on Conflict and Fragility (INCAF), and the Network on Development Evaluation (EVALNET)—in addition to a number of subgroups wherein. It is also possible to initiate informal, ad hoc groups to start discussion on a topic (e.g. the Temporary Working Group on Refugees and Migration).

Working groups meetings serve to debate issues and reach agreement on goals and standards that all member states are willing to commit to. Once consensus is reached on an issue, it can be enshrined in different kinds of “decision products.” Starting from the most formalized of these, the DAC—as any other OECD committee—can produce various types of legal instruments. These range from fully binding “international agreements” and “decisions,” although the DAC has never produced either of these kinds of products; to the non-legally binding “recommendations” and “declarations.” Official committee recommendations entail a monitoring mechanism whereby the OECD reports on each member’s compliance within five years from their adoption, while declarations are meant to set general principles or long-term goals without a requirement to monitor compliance.

In its six decades of activity, the OECD DAC has produced 16 recommendations (of which nine are currently in force), and three declarations (of which two remain in force). Well-known examples include the

2001 DAC Recommendation on Untying Official Development Assistance and the 2005 Paris Declaration on Aid Effectiveness, as well as more recent products like the 2019 DAC Recommendation on Ending Sexual Exploitation, Abuse, and Harassment in Development Co-operation and Humanitarian Assistance—a clear instance of the organization quickly responding to external events, i.e. the Oxfam scandal.

Short of enshrining a norm or best practice into one of these kinds of legal instruments in the OECD compendium, working groups can also produce decision products in the form of principles, guidelines, “lowercase-r” recommendations. Recent examples include the 2017–2019 Kampala Principles for Effective Private Sector Engagement through Development Co-operation and the 2019 report *Aligning Development Co-operation and Climate Action*. It is possible that the policies promoted in these less formal decision products are still too controversial for them to be able to find the support necessary to be enshrined into an official DAC declaration or recommendation. However, informal interviews with OECD officials have also revealed that DAC member states are weary of widening the list of official DAC legal instruments to include too many items, and prefer keeping the OECD compendium to a relatively small number of more important and comprehensive instruments.

A2 OECD DAC Peer Reviews: The Data

The analyses presented in this paper rely on metadata about the OECD DAC donor peer reviews. We collected information about the entire universe of existing peer reviews—516 reviews conducted 1962–2020—through documents available at the OECD library.¹³ For all 516 reviews, we know the DAC member who was reviewed, the date of the review, and the peer examiner countries, as we are able to retrieve this information without the need to access the full peer review report. These are the data we use in the analyses of the choice of peer examiners.

For 121 of the most recent reviews, we also have access to the full peer review report, which are available online through the OECD library.¹⁴ This includes most (but not all) reviews from the late 1990s, and all reviews starting from the year 2000. The full peer review documents contain information about the in-depth assessments of one or more recipients of the reviewee donor’s aid, which was collected manually by research assistants. They also allow us to classify reviews by document format, an indication of the different “eras” in which reviews were conducted under different operating guidelines. These are the data we use in the analyses of the choice of recipients for in-depth assessments.

Figure A1 illustrates the universe of peer reviews, highlighting: (i) the reviews for which we have access to

¹³Specifically, information about review dates and examiners for older review was retrieved from the “Chronological List of Development Assistance Committee (DAC) Aid Review (Peer Review) Meetings (by year) - 1962 to 1994.pdf” document.

¹⁴The available peer reviews are stored under slightly different classifications over time, e.g.: “Development Co-operation Reviews,” “OECD Development Assistance Peer Reviews,” and “OECD Development Co-operation Peer Reviews.”

the full report document; (ii) those for which, in spite of having access to the full document, we cannot find information on in-depth recipient assessments; and (iii) the early reviews which featured three peer examiner countries rather than two. Figure A2 focuses on the 121 reviews for which we have full documentation and illustrates the different review formats.

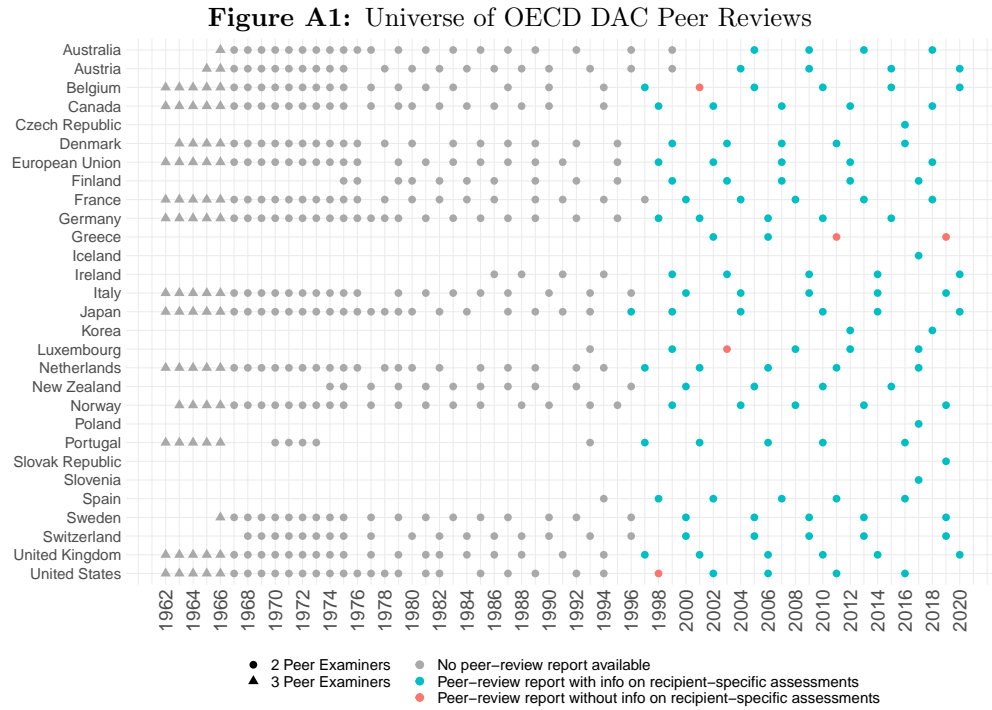
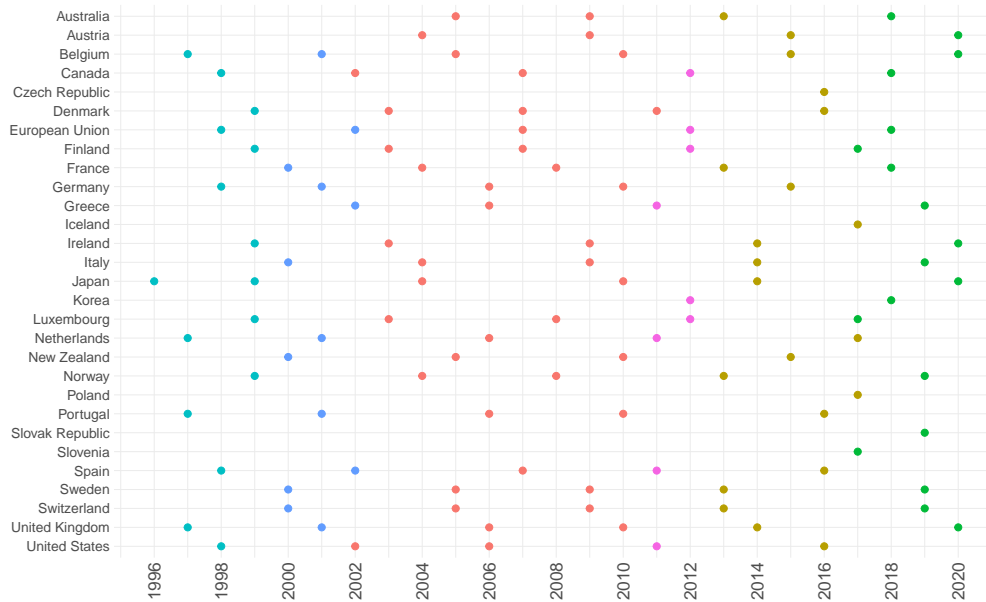


Figure A2: Distribution of Formats Across Available Peer-Review Documents



A3 Determinants of Peer Examiner Choice: The Data

Table A1: Definitions and Descriptive Statistics for Data on Peer Examiner Choice

Variable Name	Description	Count	Mean	SD	Min	Max
Is examiner	Binary indicator for whether a given pair of examiners conducted a given review (OECD 2020)	87941	0.058	0.076	0	1
At least one major donor	Binary indicator for whether at least one examiner is a major donor (United States, United Kingdom, Germany, France, Japan);	87941	0.356	0.479	0	1
Both major donors	Binary indicator for whether both examiners are major donors (United States, United Kingdom, Germany, France, Japan); the theoretically interesting effect can be obtained by adding the coefficient for “one major donor” because it includes cases of two major donors	87941	0.031	0.174	0	1
At least one similar size	Binary indicator for whether at least one reviewee-examiner pair has the same value on major donor (based on the above definition)	87941	0.826	0.379	0	1
Both similar size	Binary indicator for whether both reviewee-examiner pairs individually have identical values on major donor (based on the above definition)	87941	0.501	0.5	0	1
Average ODA budget ratio	Geometric mean of the ratio of (logged) ODA budgets between the reviewee and examiners; data on ODA commitments available from DAC1 table (OECD 2019)	80497	1.043	0.233	0.41	2.476
Dispersion of ODA budget ratio	Ratio of the larger ODA budget ratio and the smaller ODA budget ratio across both reviewee-examiner pairs; underlying data available from DAC1 table (OECD 2019)	84203	1.247	0.23	1	2.757
At least one EU member	Binary indicator for whether at least one examiner is a EU member state; this information is time-varying	87941	0.717	0.45	0	1
Both EU members	Binary indicator for whether both examiners are EU member states; the theoretically interesting effect can be obtained by adding the coefficient for “one EU member” (which includes cases of two EU members)	87941	0.242	0.428	0	1
At least one similar EU	Binary indicator for whether at least one reviewee-examiner pair has the same status with respect to being a EU member in the year of review	87491	0.701	0.458	0	1
Both similar EU	Binary indicator for whether both examiners have the same status as the reviewee with respect to being a EU member in the year of review	87491	0.247	0.431	0	1
At least one common language	Binary indicator for whether at least one examiner has a common language with the reviewee (Mayer and Zignago 2011)	87941	0.241	0.427	0	1
Both common language	Binary indicator for whether at both examiners share a common language with the reviewee (Mayer and Zignago 2011); the theoretically interesting effect can be obtained by adding the coefficient for “one examiner speaks language” (which includes cases of two examiners speaking the same language)	87941	0.028	0.165	0	1
At least one contiguous	Binary indicator for whether at least one reviewee-examiner pair entails states that share a border (Mayer and Zignago 2011)	83646	0.191	0.393	0	1
Both contiguous	Binary indicator for whether both reviewee-examiner pairs are contiguous	83646	0.026	0.158	0	1
Review twice in a row	Binary indicator for whether both examiners also reviewed the reviewee at its immediately prior review	87941	0.189	0.391	0	1
Both before as pair	Binary indicator for whether both examiners as a pair reviewed the reviewee before at any point in time	87941	0.065	0.247	0	1
At least one before	Binary indicator for whether at least one examiner was allocated this reviewee ever before (OECD 2020)	87941	0.797	0.402	0	1
Both before	Binary indicator for whether both examiners were each (individually) allocated this reviewee ever before (OECD 2020); the theoretically interesting effect can be obtained by adding the coefficient for “one examiner reviewed reviewee” (which includes cases of two examiners)	87941	0.408	0.491	0	1

Variable Name	Description	Count	Mean	SD	Min	Max
Average UNGA distance	Average of the pairwise idealpoint distances between each examiner and the reviewee, based on UN General Assembly voting behavior (Bailey, Strezhnev, and Voeten 2017)	71445	0.021	0.576	-2.444	2.216
Dispersion of UNGA distances	Dispersion of UNGA distances, defined as the difference between the larger distance and the smaller distance across both pairs; distances are absolute distances	71445	0.347	0.431	0	2.857
Average power difference	Average across both reviewee-examiner pairs of the difference in the CINC index between reviewee and examiner; CINC data available from MNC data version 5 (Singer 1972)	65171	0.002	0.05	-0.208	0.209
Dispersion in power differences	Dispersion in CINC differences, defined as the difference between the larger difference and the smaller difference across pairs	65171	0.022	0.041	0	0.208
Average partisan distance	Average of the partisan ideology distance between each examiner and the reviewee, based on a quasi-continuous partisan ideology measure (Bjornskov and Potrafke 2013)	66640	0.423	0.297	0	1.5
Dispersion of partisan distances	Dispersion of partisan ideology distances, defined as the difference between the larger distance and the smaller distance across both pairs; distances are in absolute terms	66640	0.289	0.324	0	1.5
Average difference in deliberative democracy	Average of the differences in the V-Dem deliberative democracy index (Coppedge et al. 2016) between each examiner and the reviewee; direction of difference is preserved;	73284	0.009	0.136	-0.832	0.832
Dispersion of differences in deliberative democracy	Dispersion of differences in deliberative democracy, defined as the difference between the higher difference and the lower difference in deliberative democracy across both pairs	73284	0.072	0.122	0	0.832
Arms trade between at least one	Binary indicator for whether at least one reviewee-examiner pair had any arms trade in the year of review (SIPRI 2021)	87941	0.449	0.497	0	1
Arms trade between both	Binary indicator for whether both reviewee-examiner pairs had any arms trade in the year of review (SIPRI 2021)	87941	0.116	0.32	0	1
Alliance between at least one	Binary indicator for whether at least one reviewee-examiner pair had any alliance obligation, which includes bilateral alliances and joint membership in multilateral alliances—drawn from ATOP database (Leeds et al. 2002)	87941	0.405	0.491	0	1
Alliance between both	Binary indicator for whether both reviewee-examiner pairs had alliance obligations, which includes bilateral alliances and joint membership in multilateral alliances (Leeds et al. 2002)	87941	0.145	0.352	0	1
Diplomatic ties between at least one	Binary indicator for whether at least one reviewee-examiner pair had any diplomatic representation (Bayer 2006). Since the data end in 2005, we carried them forward based on the last available year	87941	0.695	0.46	0	1
Diplomatic ties between both	Binary indicator for whether both reviewee-examiner pairs had any diplomatic representation (Bayer 2006). Since the data end in 2005, we carried them forward based on the last available year	87941	0.625	0.484	0	1
At least one same CPE	Binary indicator for whether at least one reviewee-examiner pair had the same CPE, that is, whether they were both LMEs or CMEs (Dietrich 2021)	87941	0.865	0.342	0	1
Both same CPE	Binary indicator for whether both reviewee-examiner pairs had the same CPE (Dietrich 2021)	87941	0.603	0.489	0	1
First pair role-swopped	Binary indicator for whether the current reviewee was an examiner at the most recent (past) review in which the first examiner was being reviewed (OECD 2020)	87941	0.66	0.474	0	1
Second pair role-swopped	Binary indicator for whether the current reviewee was an examiner at the most recent (past) review in which the second examiner was being reviewed (OECD 2020)	87941	0.604	0.489	0	1
Average ODA/GNI (%)	Average ODA/GNI (%) quota across both examiners, available from DAC1 table (OECD 2019)	85734	0.416	0.187	0.029	1.405
(Logged) average ODA	Average ODA budget of both examiners, available from DAC1 table (OECD 2019)	85810	7.502	0.972	3.928	10.348
Number of available examiners	Number of potential examiners available for review, which is determined by DAC membership only (OECD 2020)	87941	19.557	4.738	10	28
Three examiners	Binary indicator for whether the review involved three examiners (OECD 2020)	87941	0.06	0.237	0	1
Review year	Year of the review	87941	1991.39	17.609	1962	2020

A4 Determinants of Peer Examiner Choice: Robustness Tests

Table A2: Determinants of Peer Examiner Choice: Additional Donor Politics Variables

	(A1)	(A2)	(A3)
Arms trade between at least one	0.097 (0.061)	0.042 (0.087)	0.045 (0.085)
Arms trade between both	0.179* (0.105)	0.172 (0.126)	0.146 (0.126)
Alliance between at least one	-0.013 (0.081)	-0.018 (0.105)	0.011 (0.106)
Alliance between both	0.019 (0.089)	0.128 (0.116)	0.190 (0.117)
Diplomatic ties between at least one	-0.515 (0.384)	-0.285 (0.660)	-0.316 (0.662)
Diplomatic ties between both	-0.098 (0.113)	-0.093 (0.129)	-0.041 (0.131)
Average UNGA distance		-0.133 (0.136)	-0.013 (0.151)
Dispersion of UNGA distances		-0.099 (0.102)	-0.167 (0.105)
Average power difference		0.753 (2.177)	-0.793 (2.195)
Dispersion in power difference		1.011 (1.534)	0.898 (1.557)
Average partisan distance		0.104 (0.139)	0.106 (0.138)
Dispersion of partisan distances		0.007 (0.117)	0.002 (0.116)
Average difference in democracy		0.138 (0.597)	0.040 (0.602)
Dispersion democracy differences		-0.964** (0.399)	-0.731* (0.403)
At least one similar size			0.630*** (0.139)
Both similar size			-0.095 (0.106)
At least one similar EU			0.090 (0.092)
Both similar EU			-0.144* (0.083)
At least one common language			0.204* (0.106)
Both common language			0.313 (0.251)
At least one contiguous			-0.223** (0.105)
Both contiguous			-0.074 (0.269)
At least one same model			-0.074 (0.075)
Both same model			0.079 (0.126)
Review twice in a row			-0.196** (0.097)
Both before as pair			0.105 (0.164)
At least one before			0.004 (0.117)
Both before			0.124 (0.088)
Observations	83646	52387	52387
Within R2	0.000	0.000	0.001

Linear regressions with review fixed effects and standard errors clustered on reviews in parentheses.
Significance levels: * <.1 ** <.05 *** <.01

Table A3: Determinants of Peer Examiner Choice: Alternative Measure of Donor Size

	(A4)	(A5)	(A6)
At least one similar size (GDP-5)	0.512*** (0.087)		0.528*** (0.108)
Both similar size (GDP-5)	-0.055 (0.064)		-0.019 (0.096)
At least one similar EU	0.080 (0.069)		0.107 (0.091)
Both similar EU	-0.070 (0.067)		-0.130 (0.083)
At least one common language	0.106 (0.082)		0.211** (0.106)
Both common language	0.260 (0.198)		0.315 (0.250)
At least one contiguous	-0.056 (0.084)		-0.216** (0.104)
Both contiguous	-0.141 (0.182)		-0.080 (0.268)
At least one same model	-0.043 (0.056)		-0.100 (0.075)
Both same model	0.025 (0.096)		0.084 (0.126)
Review twice in a row	-0.064 (0.074)		-0.180* (0.098)
Both before as pair	0.162 (0.133)		0.128 (0.163)
At least one before	-0.007 (0.085)		0.005 (0.117)
Both before	0.121* (0.064)		0.122 (0.087)
Average UNGA distance		-0.157 (0.134)	-0.084 (0.139)
Dispersion of UNGA distances		-0.102 (0.101)	-0.103 (0.103)
Average power difference		0.696 (2.141)	-0.728 (2.180)
Dispersion in power difference		1.025 (1.516)	0.362 (1.570)
Average partisan distance		0.106 (0.140)	0.106 (0.139)
Dispersion of partisan distances		0.012 (0.116)	0.022 (0.116)
Average difference in democracy		0.042 (0.575)	-0.119 (0.584)
Dispersion democracy differences		-0.925** (0.388)	-0.693* (0.394)
Observations	83646	52387	52387
Within R2	0.001	0.000	0.001

Linear regression with review fixed effects and standard errors clustered on reviews in parentheses.
Significance levels: * <.1 ** <.05 *** <.01

Table A4: Determinants of Peer Examiner Choice: Alternative Measure of Donor Size

	(A7)		(A8)		(A9)	
Average ODA budget ratio	-1.121***	(0.196)			-1.916***	(0.379)
Dispersion of ODA budget ratio	0.073	(0.112)			0.411**	(0.203)
At least one similar EU	0.107	(0.069)			0.142	(0.091)
Both similar EU	-0.013	(0.068)			-0.096	(0.085)
At least one common language	0.097	(0.085)			0.254**	(0.108)
Both common language	0.220	(0.197)			0.302	(0.249)
At least one contiguous	-0.111	(0.087)			-0.276**	(0.108)
Both contiguous	-0.269	(0.176)			-0.305	(0.237)
At least one same model	0.005	(0.058)			-0.096	(0.076)
Both same model	0.037	(0.101)			0.105	(0.131)
Review twice in a row	-0.089	(0.078)			-0.132	(0.104)
Both before as pair	0.177	(0.137)			0.128	(0.165)
At least one before	-0.040	(0.093)			-0.034	(0.128)
Both before	0.013	(0.073)			0.015	(0.096)
Average UNGA distance			-0.157	(0.134)	0.007	(0.156)
Dispersion of UNGA distances			-0.102	(0.101)	-0.152	(0.111)
Average power difference			0.696	(2.141)	3.928	(2.386)
Dispersion in power difference			1.025	(1.516)	1.151	(1.659)
Average partisan distance			0.106	(0.140)	0.129	(0.142)
Dispersion of partisan distances			0.012	(0.116)	0.049	(0.118)
Average difference in democracy			0.042	(0.575)	0.341	(0.840)
Dispersion democracy differences			-0.925**	(0.388)	-0.638	(0.730)
Observations	77007		52387		49879	
Within R2	0.001		0.000		0.001	

Linear regression with review fixed effects and standard errors clustered on reviews in parentheses.
Significance levels: * <.1 ** <.05 *** <.01

Table A5: Determinants of Peer Examiner Choice: Random Effects and Additional Controls

	(A10)		(A11)		(A12)	
At least one similar size	0.360***	(0.075)			0.457***	(0.113)
Both similar size	-0.167***	(0.064)			-0.192**	(0.092)
At least one similar EU	0.063	(0.064)			0.082	(0.086)
Both similar EU	-0.071	(0.065)			-0.144*	(0.081)
At least one common language	0.064	(0.064)			0.140	(0.087)
Both common language	0.204	(0.194)			0.232	(0.245)
At least one contiguous	-0.037	(0.070)			-0.153*	(0.089)
Both contiguous	-0.120	(0.179)			-0.025	(0.267)
At least one same model	-0.009	(0.054)			-0.053	(0.072)
Both same model	0.030	(0.095)			0.092	(0.124)
Review twice in a row	-0.082	(0.074)			-0.159*	(0.095)
Both before as pair	0.172	(0.133)			0.136	(0.162)
At least one before	-0.002	(0.065)			-0.009	(0.094)
Both before	0.106*	(0.060)			0.109	(0.080)
Available examiners	-0.040	(0.082)	-0.023	(0.152)	-0.102	(0.230)
Three examiners	0.868	(1.420)	1.447	(1.745)	0.505	(2.635)
Average UNGA distance			-0.052	(0.054)	0.023	(0.062)
Dispersion of UNGA distances			-0.098	(0.100)	-0.177*	(0.104)
Average power difference			0.315	(0.582)	0.468	(0.639)
Dispersion in power difference			1.249	(1.226)	1.437	(1.317)
Average partisan distance			0.104	(0.104)	0.110	(0.103)
Dispersion of partisan distances			0.011	(0.113)	0.007	(0.113)
Average difference in democracy			-0.010	(0.337)	-0.020	(0.346)
Dispersion democracy differences			-0.870***	(0.328)	-0.690**	(0.330)
Observations	83646		52387		52387	
Within R2	0.001		0.000		0.001	

Linear regression with random effects, year effects, and standard errors clustered on reviews in parentheses.
Significance levels: * <.1 ** <.05 *** <.01

Table A6: Determinants of Peer Examiner Choice: Review Interdependence

	(A13)		(A14)		(A15)	
At least one similar size	0.549***	(0.135)			0.625***	(0.172)
Both similar size	-0.114	(0.070)			-0.111	(0.096)
At least one similar EU	0.065	(0.070)			0.092	(0.093)
Both similar EU	-0.070	(0.066)			-0.137*	(0.077)
At least one common language	0.089	(0.083)			0.192*	(0.096)
Both common language	0.252	(0.155)			0.313	(0.203)
At least one contiguous	-0.057	(0.086)			-0.209**	(0.099)
Both contiguous	-0.155	(0.168)			-0.090	(0.279)
At least one same model	-0.020	(0.064)			-0.072	(0.081)
Both same model	0.025	(0.081)			0.083	(0.114)
Review twice in a row	-0.085	(0.088)			-0.190*	(0.101)
Both before as pair	0.145	(0.130)			0.108	(0.169)
At least one before	0.001	(0.079)			0.006	(0.108)
Both before	0.125*	(0.068)			0.128	(0.086)
Average UNGA distance			-0.157	(0.131)	-0.026	(0.146)
Dispersion of UNGA distances			-0.102	(0.087)	-0.177*	(0.097)
Average power difference			0.696	(2.036)	-0.650	(2.079)
Dispersion in power difference			1.025	(1.018)	0.961	(1.142)
Average partisan distance			0.106	(0.144)	0.109	(0.143)
Dispersion of partisan distances			0.012	(0.120)	0.013	(0.117)
Average difference in democracy			0.042	(0.358)	-0.075	(0.370)
Dispersion democracy differences			-0.925***	(0.248)	-0.651**	(0.243)
Observations	83646		52387		52387	
Within R2	0.001		0.000		0.001	

Linear regression with review fixed effects and standard errors clustered on years in parentheses.

Significance levels: * <.1 ** <.05 *** <.01

Table A7: Determinants of Peer Examiner Choice: Added Review Characteristics

	(A16)		(A17)		(A18)	
Highest unemployment	-0.002	(0.003)	-0.003	(0.005)	-0.001	(0.005)
Lowest growth	-0.017	(0.017)	-0.014	(0.023)	-0.010	(0.023)
Highest corruption	-0.297	(0.263)	-0.149	(0.433)	-0.063	(0.443)
At least one similar size	0.367***	(0.140)			0.504***	(0.169)
Both similar size	-0.088	(0.072)			-0.067	(0.114)
At least one similar EU	0.197**	(0.076)			0.210**	(0.098)
Both similar EU	0.006	(0.073)			-0.015	(0.096)
At least one common language	0.025	(0.093)			0.113	(0.119)
Both common language	0.089	(0.210)			0.204	(0.265)
At least one contiguous	-0.103	(0.090)			-0.227**	(0.113)
Both contiguous	-0.020	(0.221)			-0.070	(0.282)
At least one same model	0.066	(0.061)			-0.043	(0.083)
Both same model	-0.064	(0.104)			-0.068	(0.129)
Review twice in a row	-0.004	(0.089)			-0.197*	(0.119)
Both before as pair	0.186	(0.154)			0.084	(0.179)
At least one before	0.055	(0.081)			0.162	(0.112)
Both before	0.138*	(0.074)			0.127	(0.099)
Average UNGA distance			-0.161	(0.175)	-0.024	(0.195)
Dispersion of UNGA distances			-0.215*	(0.118)	-0.288**	(0.123)
Average power difference			-0.319	(2.606)	-2.684	(2.628)
Dispersion in power difference			1.973	(1.595)	2.317	(1.632)
Average partisan distance			0.111	(0.156)	0.123	(0.153)
Dispersion of partisan distances			0.095	(0.126)	0.091	(0.125)
Average difference in democracy			0.095	(0.921)	-0.194	(0.940)
Dispersion democracy differences			-0.665	(0.657)	-0.352	(0.663)
Observations	58139		38005		38005	
Within-R2	0.001		0.001		0.001	

Linear regression with review fixed effects and standard errors clustered on reviews in parentheses.

Significance levels: * <.1 ** <.05 *** <.01

Table A8: Determinants of Peer Examiner Choice: Additional Potential Determinants

	(A19)		(A20)		(A21)	
At least one same CPE	0.014	(0.081)	-0.007	(0.085)	-0.052	(0.124)
Both same CPE	-0.020	(0.067)	0.000	(0.069)	-0.027	(0.102)
First pair role-swopped	0.212***	(0.058)	0.189***	(0.062)	0.151*	(0.082)
Second pair role-swopped	0.289***	(0.055)	0.262***	(0.061)	0.341***	(0.081)
At least one similar size			0.525***	(0.111)	0.606***	(0.139)
Both similar size			-0.112*	(0.066)	-0.116	(0.106)
At least one similar EU			0.059	(0.070)	0.079	(0.092)
Both similar EU			-0.063	(0.067)	-0.134	(0.082)
At least one common language			0.093	(0.083)	0.209*	(0.107)
Both common language			0.256	(0.199)	0.327	(0.252)
At least one contiguous			-0.056	(0.083)	-0.187*	(0.103)
Both contiguous			-0.150	(0.182)	-0.075	(0.268)
At least one same model			-0.025	(0.056)	-0.077	(0.076)
Both same model			0.020	(0.097)	0.075	(0.129)
Review twice in a row			-0.074	(0.075)	-0.179*	(0.097)
Both before as pair			0.137	(0.134)	0.099	(0.164)
At least one before			-0.056	(0.087)	-0.038	(0.120)
Both before			0.024	(0.069)	0.036	(0.091)
Average UNGA distance					0.019	(0.152)
Dispersion of UNGA distances					-0.157	(0.106)
Average power difference					-0.656	(2.152)
Dispersion in power difference					0.833	(1.545)
Average partisan distance					0.097	(0.138)
Dispersion of partisan distances					0.022	(0.117)
Average difference in democracy					-0.003	(0.583)
Dispersion democracy differences					-0.620	(0.394)
Observations	83646		83646		52387	
Within R2	0.000		0.001		0.001	

Linear regression with review-fixed effects and standard errors clustered on reviews in parentheses.
Significance levels: * <.1 ** <.05 *** <.01

A5 Determinants of Recipient Choice: The Data

A5.1 Data on Official Development Assistance

The data on foreign aid flows used in this paper come from the OECD DAC Creditor Reporting System (CRS). The version of these data used in this draft was updated on the OECD.Stat website on April 28 and downloaded by us on May 6, 2021.¹⁵ We aggregate the CRS’s project-level information into yearly figures by donor, recipient, and donor-recipient pairs as necessary. In our aggregation, we only consider flows marked as “ODA Grants” or “ODA Loans”—thus excluding flows marked as “Equity Investment,” “Private Development Finance,” and “Other Official Flows (non Export Credit)” from our analyses. The resulting

¹⁵<https://stats.oecd.org/DownloadFiles.aspx?DatasetCode=CRS1>

data includes information on 47 years, from 1973 to 2019; 95 donors;¹⁶ and 203 recipients¹⁷ (note that 21 of the recipients are “regional”¹⁸ and two are “unspecified”¹⁹).

We use commitment data in million US dollars, 2019 constant amounts (variable `USD_Commitment_Def1`, which for 2019 data is the same as variable `USD_Commitment`). Note that, out of the 3,986,771 project-level observations from 1990 to 2019, 265 have a negative commitment amount (0.0066%); 802,917 have a commitment amount equal to zero (20%); and 1,013,029 are listed as not available (25%). As we aggregate amounts from the project level to construct donor-year, recipient-year, and donor-recipient-year totals, we sum across all projects—including those with zero or negative amounts—while the projects with no commitment amount information are dropped.

¹⁶Adaptation Fund; African Development Bank; African Development Fund; Arab Bank for Economic Development in Africa; Arab Fund (AFESD); Asian Development Bank; Asian Infrastructure Investment Bank; Australia; Austria; Azerbaijan; Belgium; Bulgaria; Canada; Caribbean Development Bank; Center of Excellence in Finance; Central Emergency Response Fund; Chinese Taipei; Climate Investment Funds; Council of Europe Development Bank; Croatia; Cyprus; Czech Republic; Denmark; Development Bank of Latin America; Estonia; EU Institutions; Finland; Food and Agriculture Organisation; France; Germany; Global Alliance for Vaccines and Immunization; Global Environment Facility; Global Fund; Global Green Growth Institute; Greece; Green Climate Fund; Hungary; Iceland; IFAD; IMF (Concessional Trust Funds); Inter-American Development Bank; International Atomic Energy Agency; International Bank for Reconstruction and Development; International Development Association; International Labour Organisation; Ireland; Islamic Development Bank; Israel; Italy; Japan; Kazakhstan; Korea; Kuwait; Latvia; Liechtenstein; Lithuania; Luxembourg; Malta; Montreal Protocol; Netherlands; New Zealand; Nordic Development Fund; Norway; OPEC Fund for International Development; OSCE; Poland; Portugal; Qatar; Romania; Russia; Saudi Arabia; Slovak Republic; Slovenia; Spain; Sweden; Switzerland; Thailand; Timor-Leste; Turkey; UN Institute for Disarmament Research; UN Peacebuilding Fund; UNAIDS; UNDP; UNECE; UNEP; UNFPA; UNHCR; UNICEF; United Arab Emirates; United Kingdom; United States; UNRWA; WFP; World Health Organisation; World Tourism Organisation.

¹⁷Afghanistan; Africa, regional; Albania; Algeria; America, regional; Angola; Anguilla; Antigua and Barbuda; Argentina; Armenia; Aruba; Asia, regional; Azerbaijan; Bahamas; Bahrain; Bangladesh; Barbados; Belarus; Belize; Benin; Bermuda; Bhutan; Bilateral, unspecified; Bolivia; Bosnia and Herzegovina; Botswana; Brazil; British Virgin Islands; Brunei Darussalam; Burkina Faso; Burundi; Cabo Verde; Cambodia; Cameroon; Caribbean Central America, regional; Caribbean, regional; Cayman Islands; Central African Republic; Central America, regional; Central Asia, regional; Chad; Chile; China (People’s Republic of); Chinese Taipei; Colombia; Comoros; Congo; Cook Islands; Costa Rica; Croatia; Cuba; Cyprus; Cote d’Ivoire; Democratic People’s Republic of Korea; Democratic Republic of the Congo; Djibouti; Dominica; Dominican Republic; East African Community; Eastern Africa, regional; Ecuador; Egypt; El Salvador; Equatorial Guinea; Eritrea; Eswatini; Ethiopia; Europe, regional; Far East Asia, regional; Fiji; French Polynesia; Gabon; Gambia; Georgia; Ghana; Gibraltar; Grenada; Guatemala; Guinea; Guinea-Bissau; Guyana; Haiti; Honduras; Hong Kong (China); India; Indonesia; Iran; Iraq; Israel; Jamaica; Jordan; Kazakhstan; Kenya; Kiribati; Korea; Kosovo; Kuwait; Kyrgyzstan; Lao People’s Democratic Republic; Lebanon; Lesotho; Liberia; Libya; Macau (China); Madagascar; Malawi; Malaysia; Maldives; Mali; Malta; Marshall Islands; Mauritania; Mauritius; Mayotte; Melanesia, regional; Mexico; Micronesia; Middle Africa, regional; Middle East, regional; Moldova; Mongolia; Montenegro; Montserrat; Morocco; Mozambique; Myanmar; Namibia; Nauru; Nepal; Netherlands Antilles; New Caledonia; Nicaragua; Niger; Nigeria; Niue; North Macedonia; North of Sahara, regional; Northern Mariana Islands; Oceania, regional; Oman; Pakistan; Palau; Panama; Papua New Guinea; Paraguay; Peru; Philippines; Qatar; Rwanda; Saint Helena; Saint Kitts and Nevis; Saint Lucia; Saint Vincent and the Grenadines; Samoa; Sao Tome and Principe; Saudi Arabia; Senegal; Serbia; Seychelles; Sierra Leone; Singapore; Slovenia; Solomon Islands; Somalia; South Central Asia, regional; South Africa; South America, regional; South Asia, regional; South of Sahara, regional; South Sudan; Southern Africa, regional; Sri Lanka; States Ex-Yugoslavia unspecified; Sudan; Suriname; Syrian Arab Republic; Tajikistan; Tanzania; Thailand; Timor-Leste; Togo; Tokelau; Tonga; Trinidad and Tobago; Tunisia; Turkey; Turkmenistan; Turks and Caicos Islands; Tuvalu; Uganda; Ukraine; United Arab Emirates; Uruguay; Uzbekistan; Vanuatu; Venezuela; Viet Nam; Wallis and Futuna; West Bank and Gaza Strip; Western Africa, regional; Yemen; Zambia; Zimbabwe.

¹⁸Africa, regional; America, regional; Asia, regional; Caribbean Central America, regional; Caribbean, regional; Central America, regional; Central Asia, regional; Eastern Africa, regional; Europe, regional; Far East Asia, regional; Melanesia, regional; Middle Africa, regional; Middle East, regional; North of Sahara, regional; Oceania, regional; South Central Asia, regional; South America, regional; South Asia, regional; South of Sahara, regional; Southern Africa, regional; Western Africa, regional.

¹⁹Bilateral, unspecified; States Ex-Yugoslavia unspecified.

A5.2 Data on Recipient Characteristics

A5.2.1 Quality of Government (QoG) Data

Many of the recipient characteristics included in our models are data from the Quality of Government (QoG) Basic Dataset 2021.²⁰ Specifically, we use the tenfold classification of colonial origin from Wahman, Teorell and Hadenius (Hadenius and Teorell, 2007; Teorell and Wahman, 2018; Wahman et al., 2013)—variable `ht_colonial` in the QoG Basic Dataset 2021—to construct our indicator of colonial history, which equal to one for recipients who were primarily a former colony of the donor, and zero otherwise.

We measure democracy with the Revised Combined Polity Score (`p_polity2`) from the Polity project (Marshall and Gurr, 2020); and government effectiveness (`wbgi_gee`), rule of law (`wbgi_rle`), and political stability and absence of violence/terrorism (`wbgi_pve`) with estimates from the World Bank’s Governance Indicators. We also use measures of life expectancy at birth (`wdi_lifexp`) and GDP per capita in constant 2010 US dollars (`wdi_gdpcapcon2010`) from the Bank’s World Development Indicators.

Note that, when including these covariates, we lose observations from the 22 recipients who are not countries in the QoG data.²¹ This is a problem for Kosovo, which was visited once by Austria in 2020; and for the West Bank and Gaza Strip, which was visited by Italy in 2000. Since Italy also visited Ethiopia in 2000, that review remains in the sample. However, Austria only visited Kosovo in 2020—this means that lines from the Austria 2020 review remain in the sample for our models, but no recipient is selected with a value of one for the dependent variable once the Austria-2020-Kosovo observation is dropped.

A5.2.2 UN Security Council Data

We include an indicator for whether a recipient country holds one of the rotating positions in the UN Security Council in the year of a review. These data are from Dreher, Sturm, and Vreeland (2009) and were updated by Dreher, Lang, Rosendorff, and Vreeland (2018).²²

Note that the following recipients do not appear in these data: Anguilla, Aruba, Bermuda, British Virgin Islands, Cayman Islands, China (People’s Republic of), Cook Islands, French Polynesia, Gibraltar, Hong Kong (China), Kosovo, Macau (China), Mayotte, Montenegro, Montserrat, Netherlands Antilles, New Caledonia, Niue, Northern Mariana Islands, Saint Helena, Serbia, South Sudan, Timor-Leste, Tokelau, Turks and Caicos Islands, Tuvalu, Wallis and Futuna, West Bank and Gaza Strip. We manually code this indicator to be equal to one for China in all years—since it is a permanent member of the UNSC.

²⁰Retrieved on May 3, 2021, from <http://www.qog.pol.gu.se/doi:10.18157/qogbasjan21>.

²¹Anguilla, Aruba, Bermuda, British Virgin Islands, Cayman Islands, Cook Islands, French Polynesia, Gibraltar, Hong Kong (China), Kosovo, Macau (China), Mayotte, Montserrat, Netherlands Antilles, New Caledonia, Niue, Northern Mariana Islands, Saint Helena, Tokelau, Turks and Caicos Islands, Wallis and Futuna, West Bank and Gaza Strip.

²²The data were last updated on August 4, 2020, and retrieved by us on May 10, 2021, from https://www.uni-heidelberg.de/fakultaeten/wiso/awi/professuren/intwipol/datasets_en.html.

This means that, when including this indicator, we lose relevant observations for recipients who were visited in the cases of Kosovo and West Bank and Gaza Strip (already discussed); but also Montenegro and Timor-Leste.

A5.2.3 World Bank’s World Development Indicators Data

We include information on recipients’ tourism and statistical capacity from the World Bank’s World Development Indicators (WDIs). To measure a recipient’s level of tourism, we use the number of tourists who travel to a country other than that in which they have their usual residence, but outside their usual environment, for a period not exceeding 12 months and whose main purpose in visiting is other than an activity remunerated from within the country visited (“International inbound tourists (overnight visitors)” indicator). These data are originally from the *Yearbook of Tourism Statistics* of the World Tourism Organization, and they were last updated on June 30, 2021.

To measure statistical capacity in the recipient country, we use the overall average Statistical Capacity score from the World Bank’s *Bulletin Board on Statistical Capacity*, last updated on July 30, 2021. This indicator is a composite score assessing the capacity of a country’s statistical system. It is based on a diagnostic framework assessing the following areas: methodology; data sources; and periodicity and timeliness. Countries are scored against 25 criteria in these areas, using publicly available information and/or country input. The overall Statistical Capacity score is then calculated as a simple average of all three area scores on a scale of 0–100.

A5.3 Data on Recipient Assessments

We retrieved information on recipient-specific assessments of the donor’s development assistance program from the peer review reports. As depicted in Figure A1, we have access to the full report document for 121 peer reviews, 1996–2020; and 116 of them mention one or more recipient-specific assessment. The following figures provide additional descriptive information about these data.

Figure A3: Number of Recipient-Specific Assessments per Review, by Type



Figure A4: Universe of Recipient-Specific Assessments

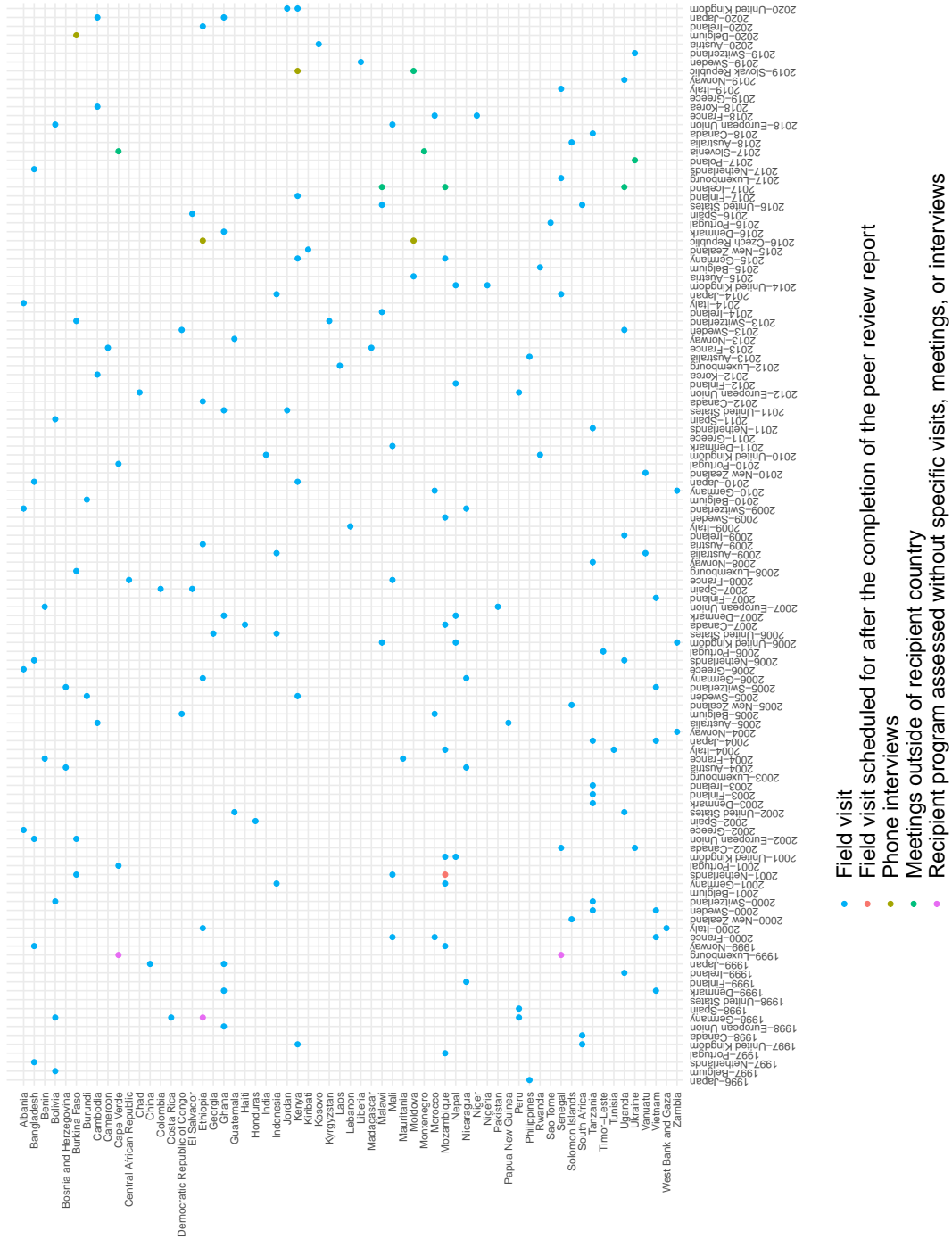


Figure A5: Number of Recipient-Specific Assessments per Donor

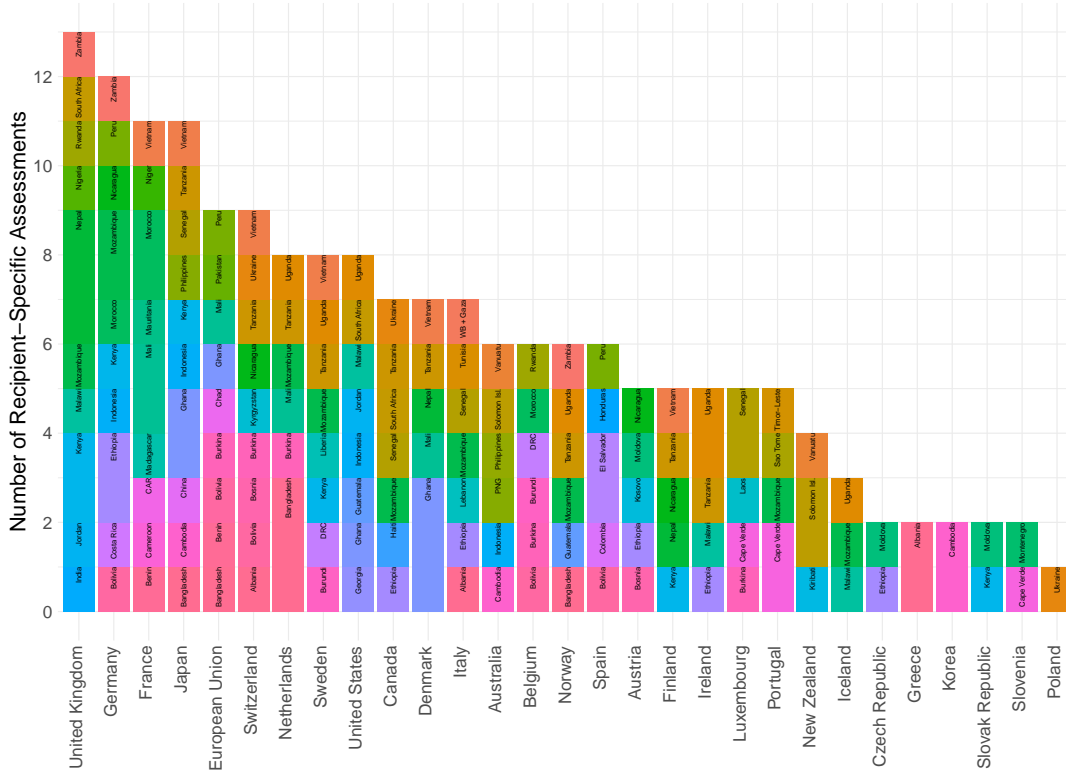
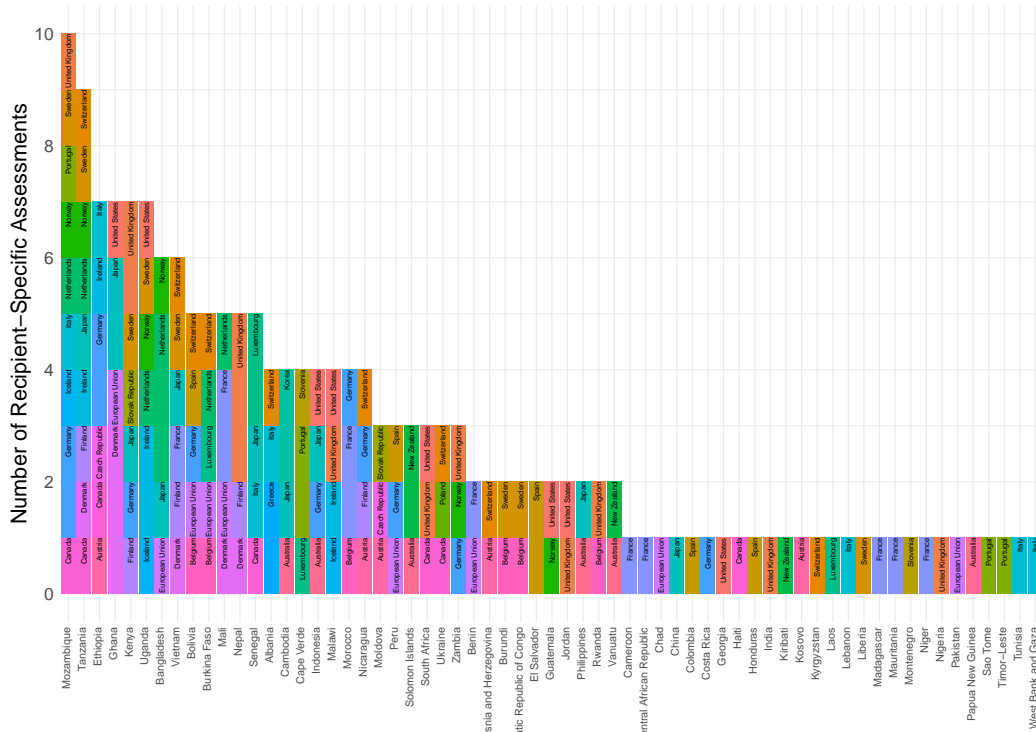


Figure A6: Number of Recipient-Specific Assessments per Recipient



A6 Determinants of Recipient Choice: Robustness Tests

The results presented in the main paper remain largely unchanged if we limit our sample to reviews of traditional donors only—meaning those donors who joined the OECD DAC before 2000 and thus excluding South Korea, Iceland, and the East European donors who joined in the 2010s.

Table A9: Determinants of Recipient Choice, Experienced Donors Only

	(A22)	(A23)	(A24)
Aid from Donor, 3-yr Avg.	0.000014 (0.000018)	0.000008 (0.000018)	0.000007 (0.000018)
Aid from Donor, % of Donor Tot.	0.011016*** (0.001441)	0.011291*** (0.001463)	0.010890*** (0.001508)
Aid from all DAC Donors, % of DAC Donors Tot.	0.004737 (0.002781)	0.004569 (0.002785)	0.011024** (0.004149)
Former Colony of Donor	0.042150*** (0.009155)	0.041055*** (0.009191)	0.042169*** (0.009273)
Democracy, Prev. Year	0.000288 (0.000371)	0.000297 (0.000371)	0.000313 (0.001073)
Government Effectiveness, Prev. Year	0.001640 (0.007709)	0.001870 (0.007753)	0.005892 (0.014286)
Rule of Law, Prev. Year	0.009517 (0.007255)	0.009805 (0.007278)	0.003445 (0.016485)
Political Stability, Prev. Year	0.004869 (0.003155)	0.005086 (0.003164)	0.004703 (0.006401)
Life expectancy at birth, Prev. Year	0.000076 (0.000277)	0.000046 (0.000283)	0.000000 (0.001212)
GDP per capita, Prev. Year	0.000003*** (0.000001)	0.000003*** (0.000001)	0.000002 (0.000004)
UNSC Membership	0.007025 (0.007012)	0.006956 (0.007018)	0.004287 (0.007579)
International tourism, Prev. Year	0.000000 (0.000000)	0.000000 (0.000000)	0.000000 (0.000000)
Statistical capacity, Prev. Year	0.000298 (0.000172)	0.000303 (0.000172)	0.000007 (0.000366)
Effects	Random	Random	Recipient FEs
Controls	None	Doc Format and Pool Size	Doc Format and Pool Size
Num.Obs.	4,775	4,775	4,775
R2	0.035	0.036	0.027
R2 Adj.	0.032	0.032	0.004

Linear regressions, standard errors in parentheses. Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001.