Cultural insiders and foreign aid: How the cultural background of World Bank project managers affects project success

Joe Mitchell-Nelson

October 2020

Abstract

This research explores how the cultural background of project leaders affects the success of foreign aid projects. I use a new measure of cultural proximity between countries, based on the genetic distance measure compiled by Spolaore and Wacziarg (2018) and data from the World Bank, to quantify how much cultural overlap likely exists between project leaders and the countries where these projects take place. I then present a principal-agent model that predicts the effort levels of project leaders as a function of cultural proximity and institutional quality. I find that this structural model describes the data better than a number of intuitive reduced-form specifications. To address possible endogeneity arising from assignment of managers to projects, I instrument for cultural proximity with the average cultural proximity of other available project leaders. Where institutions are strong, culturally similar project managers outperform those who are more culturally distant, but this relationship is not present in countries with poor institutions.

1 Introduction

Foreign aid projects can face serious and often unforeseeable challenges. Goods and services provided by foreign aid projects do not always reach recipients, often because of cost overruns and delays. The Soviet Union, for example, assisted Nigeria in building an $8 billion steel mill starting in 1979. Despite intermittent efforts to complete the mill over the last 40 years, Ajaokuta Steel Mill has never produced any iron or steel (Mold, 2012). When goods and services do find their way into the right hands, the actual recipients may find little use for them. Point-of-use water sanitation products (e.g. chlorine tablets and ceramic filters) have notoriously low usage rates, even when they are subsidized or free and come with in-person education about the dangers of untreated drinking water (Luoto et al., 2011). Finally, much-needed aid can also have unintended consequences. A World Bank project in Kenya aimed at connecting poor residents to the public electrical grid has fueled the rise of electricity cartels in some districts of Nairobi. Cartels cut the metered connections to households and then offer to shunt those same households back into the grid for a fee (Langat, 2019). Cassen (1994) reports that about one-quarter to one-third of foreign aid projects, overall, are not successful. Given that net worldwide foreign aid has climbed to over $160B annually, factors that can lower the failure rate of foreign aid projects have the potential to avert tremendous losses of scarce resources.[1]

To this end, I identify the causal effect of one potentially important but as-yet unexplored factor in the success of foreign aid projects: the cultural background of project leaders. Specifically, I construct a measure of the likely cultural proximity between recipient countries and the individuals who are assigned to supervise the planning and execution of a project. As will be discussed in more detail in Section 2, several previous studies of foreign aid projects have concluded that the identities of project supervisors matter, but none of these studies has identified exactly which characteristics of these project supervisors are most important. A project supervisor's cultural proximity, or familiarity with the culture of the recipient country,

[1]This measure of foreign aid includes only official development assistance (ODA). The OECD counts a transfer as ODA if (1) it comes from an official body (usually a government or a multilateral donor); (2) it "is administered with the promotion of the economic development and welfare of developing countries as its main objective"; and (3) it "is concessional in character and conveys a grant element of at least 25 per cent (calculated at a rate of discount of 10 percent)."

may help that supervisor produce more effective outcomes. An understanding of the cultural complexities of a recipient population seems an obvious qualification for determining and supplying the needs of that population. Cultural proximity may lead project supervisors to ask the right questions when planning the project, to catch likely problems with implementation before they become apparent to others, and to rightly trust their gut. On the other hand, I find that World Bank project supervisors, officially called task team leaders (TTLs), who are likely to be culturally similar to their recipient country have, on average, better project outcomes, provided that the recipient country has sufficiently strong institutions, as measured by the World Bank's Country Policy and Institutional Assessment (CPIA). When projects are located in countries with poor institutions, however, culturally close TTLs seem to have no advantage over their culturally more-distant counterparts. These results have clear implications for staffing and hiring decisions at multilateral aid organizations.

I do not observe each TTL's cultural backgrounds directly. Instead, I rely on *genetic distance*, an existing measure of cultural divergence between countries developed by Spolaore and Wacziarg (2009; 2016; 2018). Given that cultural traits are transmitted intergenerationally, as are genetic traits, Spolaore and Wacziarg argue that the genetic distance between populations constitutes an appropriate proxy for the cultural distance between populations.[2] The most recent measure of genetic distance developed by Spolaore and Wacziarg—the data I use here—is based on neutral genetic features that geneticists believe are not affected by selection pressures. As a result, any variation in these features across different populations can be attributed to random drift. The longer any two populations have been spatially separated, the more drift will have occurred, and the more genetically distant these two populations will be. Like these genetic traits, we expect cultural traits also to diverge further as the duration of separation increases.

Though genetic distance is conceptually a measure that compares two populations, here I use it to compare individuals to populations—namely, the TTLs who oversee aid projects on behalf of the World Bank in relation to the countries that receive this aid. In doing so, I assume that TTLs are familiar through experience with the culture of their own ancestral country. Of course, I do not assume that TTLs personally adopt every aspect of their home

[2]Genetic distance is a measure based on the similarity of the distributions of genetic markers between two populations. Appendix B provides a detailed description of how Spolaore and Wacziarg's measure of genetic distance is calculated.

country's culture, which may in fact be a patchwork of regional subcultures. Throughout this paper, I refer to the *negative* of the measured genetic distance between a TTL's ancestral country and the recipient country where the TTL supervises a foreign aid project as the *cultural proximity* between that TTL and the country where their assigned project is located.

A vast literature has examined the effectiveness of foreign aid. One branch of this literature has sought to measure the effect of foreign aid on GDP growth and other broad macroeconomic measures of development, but has arrived at no consensus about the effect of foreign aid on GDP growth.[3] In a thorough review of the literature to that date, Temple (2010) writes that the "cross-country evidence on the effects of aid must be regarded as a work-in-progress." In a more-recent review of the literature, Qian (2015) similarly argues that the effect of foreign aid on a developing country's economic growth "is perhaps among the most controversial [issues] in development and growth economics."

Another branch of the literature on the effectiveness of foreign aid focuses on the outcomes of *individual* aid projects, rather than country-level aggregate measures of success. This body of research implicitly sets a much looser criterion for the success of aid. Aid is recorded as successful when projects demonstrably and efficiently meet their narrowly defined objectives. These projects may or may not contribute to improved growth rates. Furthermore, this literature is rarely able to show that aid does not simply substitute for government expenditures, which may then be diverted to less socially desirable objectives. Also, there may be complementarities among aid projects that are not captured by project-level analyses. Nevertheless, a foreign aid project can hardly be considered successful if it fails even on its own terms.

Much of the literature concerning the success of individual aid projects has focused on characteristics of recipient countries, and of the foreign aid projects themselves, as explanatory variables (see for instance Denizer et al., 2013; Duponchel et al., 2010). For several reasons, many of these studies draw their samples from the same World Bank database of aid projects. First, the World Bank is a multilateral donor, and so is considered less likely to direct aid flows specifically to advance any single country's foreign policy objectives. Second,

[3] Some scholars find evidence that foreign aid can lift countries out of poverty (e.g. Arndt et al., 2010; Burnside and Dollar, 2000). Others report that aid does not improve growth, even in countries with strong institutions (e.g. Easterly, 2003; Doucouliagos and Paldam, 2009; Doucouliagos and Paldam, 2011; Rajan and Subramanian, 2008).

the World Bank is among the largest multilateral donors. In the last decade, the World Bank distributed more foreign aid than the United Nations, the International Monetary Fund and the World Health Organization combined. Finally, every World Bank project is uniformly assessed by the Independent Evaluation Group (IEG), a distinct arm of the World Bank, whose sole responsibility is objective evaluation of the development effectiveness of the World Bank group. For these reasons, I rely on the same World Bank data as Denizer et al. (2013) and Duponchel et al. (2010) for the analysis of foreign aid projects undertaken in this study.

Section 2 discusses some related research concerning the determinants of successful foreign aid projects. In Section 3, I outline a model of TTL effort as a function of cultural proximity and institutional quality, and derive predictions for the joint effects on project success of cultural proximity and institutional quality. Section 4 provides details about the empirical strategy I use to (1) measure cultural proximity between the TTL and the recipient country and (2) estimate the effects on project success of cultural proximity and the recipient country's institutional quality. One contribution of this paper is to combine Spolaore and Wacziarg's data with other sources to develop a plausible measure of cultural proximity between individuals and countries that does not rely on demographic data about these individuals. Section 5 presents the main results and some robustness checks. Section 6 explores and comments upon one potential source of mismeasurement in my construction of the *cultural proximity* variable, and section 7 concludes.

2 Related literature

The development economics literature has produced a number of studies that examine factors that improve the outcomes of donor-funded development projects (Dollar and Svensson, 2000; Kilby, 2000; Chauvet et al., 2017), including many studies published by the World Bank itself (Guillaumont and Laajaj, 2006; Duponchel et al., 2010).[4] While other papers in the

[4]Scholars of project management have also given some attention to development projects (Diallo and Thuillier, 2005; Ika et al., 2012). The disciplines of economics and project management bring different assumptions and methodological procedures to the question of development project effectiveness. Where the economics literature emphasizes the role of recipient country and project characteristics, the project management literature focuses on the structure of working groups and characteristics of team members and their relationships with each other.

economics literature have recognized that TTL characteristics matter, the present paper describes what seems to be the first effort to examine one specific characteristic of TTLs beyond just broad measures of their ability.

Studies published by the World Bank frequently identify recipient country institutions as one of the most important determinants of project success. Dollar and Levin (2005) use an instrumental variables strategy to determine the effect of institutional quality on development project outcomes, and find evidence of a causal, positive relationship between institutional quality and the proportion of projects in each country rated as "successful" by the World Bank's Operations Evaluation Department.[5] Geli, Kraay and Nobakht (2014) develop a relatively simple predictive model to determine World Bank project outcomes. They estimate a probit model to predict project success, using only a handful of variables: log of project cost, preparation time, initially planned project length, the outcome of TTLs' other projects, and the recipient country's institutional quality. They find that institutional quality and TTL track record are the strongest predictors of project success.

Denizer, Kaufmann and Kraay (2013) compare the effects on project outcomes of (a) country characteristics (e.g. institutions, GDP growth), and (b) project characteristics (e.g. duration, cost, sector), and find that both matter. Most relevant to the present analysis, these authors use the IEG scores that a given TTL has received on projects *other than* the current project to measure TTL quality. TTL quality is significantly and positively associated with project outcomes, but they do not address which observable TTL characteristics measure TTL quality. Furthermore, it is possible that the results in their paper may be driven by the TTL assignment process. I make two improvements: I examine a particular characteristic of each TTL—their shared cultural background with the recipient country—and I use a plausibly exogenous instrument for TTL assignment.

3 Conceptual model

TTL effort is likely to be an important factor in foreign aid project quality. Because I cannot observe effort directly, I introduce a principal-agent model to explain variation in TTL effort

[5]The Operations Evaluation Department was the precursor to the IEG, whose ratings I use to determine project quality.

as a function of the TTL's cultural proximity to the recipient country and the institutional quality of the recipient country. The World Bank provides a unique context for a principal-agent model, as TTLs are not remunerated on the basis of project quality. Instead I assume that TTLs receive professional and social benefits based on the level of effort inferred from their supervisors and peers, based on their project's quality.[6] The model begins with the following assumptions:

1. TTLs are utility maximizers who dislike effort.
2. A TTL's effort is more effective when their cultural proximity is high.
3. A project's quality is more predictable when the recipient country's institutional quality is high.
4. TTL effort is not directly observable by the principal.
5. TTLs are rewarded in proportion the probability, from the principal's perspective, that the TTL's effort exceeded some threshold.

3.1 Setting up the model

The agent chooses an effort level e to maximize the indirect utility function:

$$v(e) = R(e, \eta) - \phi(e) \tag{1}$$

where R is the reward the agent receives from the principal, η is an error term (to be described shortly), and $\phi(e)$ gives the agent's disutility of effort. R is increasing in e and η. Assume also that $\phi'(e) > 0$ and $\phi''(e) > 0$ so that marginal *dis*utility of effort is *increasing*.

The principal observes project quality Q and forms a belief about the distribution of possible levels of the TTL's true effort e. Project quality Q is determined by the following production function:

$$Q = q(X) + g(C)e + \eta \tag{2}$$

[6] These benefits may include access to more desirable projects or positions in the future, professional recognition from peers, etc.

where $q(X)$ is a function of variables X that affect project quality but are outside of the TTL's control, and $g(C)e$ is the total contribution of TTL effort to project quality. The function $g(C) > 0$ is increasing and gives the marginal contribution of effort to project quality, reflecting the assumption that TTL effort is more effective when their cultural proximity is high. The stochastic error term η is distributed with mean $\mu_\eta = 0$ and variance $\sigma_\eta^2 = h(I)$. The variance $h(I)$ of η is a decreasing function of I, reflecting the assumption that project outcomes are more predictable in recipient countries with high institutional quality I. Let $F(x; h(I))$ and $f(x; h(I))$ be, respectively, the cumulative density function and the probability density function of η.

3.2 Solving the model

The principal knows the values of Q, and X (note that $C, I \in X$), along with the functional forms of Q, q, g and h, but is unable to observe η or e. The principal can then calculate an "observed" effort level $e_o = e + \frac{\eta}{g(C)}$ by rearranging the production function:

$$Q = q(X) + g(C)e + \eta$$
$$\frac{Q - q(X)}{g(C)} = e + \frac{\eta}{g(C)} = e_o$$

Note that $E(e_o) = e$. Recall that the principal rewards the TTL in proportion to the probability that TTL effort was above some threshold value m. How does the principal use e_o to determine the probability that the agent's effort e was above the threshold m? Intuitively, if the principal observes a low level of e_o, such that $e_o < m$, they must weigh the possibility that in fact $e > m$ but that η was negative and large in magnitude—the agent may have worked hard but been unlucky. Conversely, if the principal observes $e_o > m$, it remains possible that $e < m$ but the agent was unlucky and η was positive and large. For simplicity I suppose the principal has no prior belief about the agent's likely level of effort.

Given the principal's information set, the probability that $e > m$ is

$$P(e > m) = P\left[\eta < (e_o - m)g(C)\right] = F[(e_o - m)g(C); h(I)]$$

The TTL chooses their true effort level e before observing η. From the TTL's point of view, then, the reward R can be expressed as

$$\begin{aligned} R(e,\eta) &= F[(e_o - m)g(C); h(I)] \\ &= F[(e + \frac{\eta}{g(C)} - m)g(C); h(I)] \\ &= F[(e - m)g(C) + \eta; h(I)] \end{aligned}$$

The TTL then integrates over the set of rewards times their probability to calculate the *expected* reward $E(R(e))$ for any level of effort e:

$$E(R(e)) = \int_{-\infty}^{\infty} F[g(C)(e-m) + x; h(I)]f(x; h(I))dx$$

Then the agent's expected utility is given by

$$E(v(e)) = \int_{-\infty}^{\infty} F[g(C)(e-m) + x; h(I)]f(x; h(I))dx - \phi(e)$$

which yields the following necessary condition for a maximum:

$$\frac{\partial}{\partial e}\left(\int_{-\infty}^{\infty} F[g(C)(e-m) + x; h(I)]f(x; h(I))dx\right)\bigg|_{e=e^*} - \phi'(e^*) = 0$$

Unfortunately, it is not possible to be more specific about the solution to the model without choosing an explicit distributional family for η. To proceed, I assume a logistic distribution for η.[7] Given that η has mean 0 and variance $h(I)$, we have location parameter $\mu = 0$ and scale parameter $s = \frac{\sqrt{3h(I)}}{\pi}$. For notational parsimony, also let $a = g(C)(e-m)$. The TTL's first order condition is then:

$$\frac{\partial}{\partial e}\left(\int_{-\infty}^{\infty} \frac{1}{1+\exp\left(\frac{-x-a}{s}\right)} \times \frac{\exp\left(\frac{-x}{s}\right)}{s\left(\exp\left(\frac{-x}{s}\right)+1\right)^2}dx\right)\Bigg|_{e=e^*} - \phi'(e^*) = 0$$

Solving the integral above yields the antiderivative, denoted $B(x)$:

$$B(x) = -\frac{\exp(\frac{a}{s})\left\{\left[\exp(\frac{x}{s})+1\right]\left[\ln\left(\exp(\frac{x+a}{s})+1\right) - \ln\left(\exp(\frac{x}{s})+1\right)\right] + \exp(\frac{a}{s}) - 1\right\}}{\left(\exp(\frac{a}{s})-1\right)^2\left(\exp(\frac{x}{s})+1\right)}$$

[7] I intend to include solutions for other distributional assumptions in an online appendix.

Then we take the limits of $B(x)$ as $x \to \infty$ and as $x \to -\infty$ to evaluate the integral:

$$\lim_{x\to\infty} B(x) = \frac{-a \times \exp\left(\frac{a}{s}\right)}{s\left(\exp(\frac{a}{s}) - 1\right)^2}$$
$$\lim_{x\to-\infty} B(x) = \frac{-\exp\left(\frac{a}{s}\right)}{\exp\left(\frac{a}{s}\right) - 1}$$

Subtracting $\lim_{x\to\infty} B(x) - \lim_{x\to-\infty} B(x)$ produces the following definite integral:

$$\frac{\exp(\frac{a}{s})}{\exp(\frac{a}{s}) - 1} - \frac{a \times \exp(\frac{a}{s})}{s\left(\exp(\frac{a}{s}) - 1\right)^2}$$

We can then express the TTL's optimization problem as:

$$\max_{e} \quad \frac{\exp(\frac{a}{s})}{\exp(\frac{a}{s}) - 1} - \frac{a \times \exp(\frac{a}{s})}{s\left(\exp(\frac{a}{s}) - 1\right)^2} - \phi(e)$$

Let e^* be the TTL's utility-maximizing level of effort, and let $a^* = g(C)(e^* - m)$. Then the TTL's first order condition is:

$$g(C) \times \frac{\exp(\frac{a^*}{s})\left((a^* - 2s)\exp(\frac{a^*}{s}) + a^* + 2s\right)}{s^2\left(\exp(\frac{a^*}{s}) - 1\right)^3} - \phi'(e^*) = 0 \tag{3}$$

We are ultimately interested in how the TTL's chosen effort level e^* varies with cultural proximity C and institutional quality I. That is, we would like to say something about $\frac{\partial e^*}{\partial C}$, $\frac{\partial e^*}{\partial I}$. To find and sign these partial derivatives, we turn to the implicit function theorem. Let $D(e^*, I, C)$ denote the left-hand side of equation 3.[8] Then we have:

$$\frac{\partial e^*}{\partial C} = -\frac{D_C}{D_{e^*}} \tag{4}$$
$$\frac{\partial e^*}{\partial I} = -\frac{D_I}{D_{e^*}} \tag{5}$$

where the components take the form:

[8]It seems prudent to note again here that the logistic distribution's shape parameter $s = \frac{\sqrt{3h(I)}}{\pi}$, a parameterization chosen so that $\text{var}(\eta) = h(I)$.

$$
\begin{aligned}
D_{e^*} &= -\frac{\exp(\frac{a^*}{s})\left((a^*-3s)\exp(\frac{2a^*}{s})+4a^*\exp(\frac{a^*}{s})+a^*+3s\right)}{s^3\left(\exp(\frac{a^*}{s})-1\right)^4}g(C)\frac{\partial a^*}{\partial e^*}-\phi''(e^*)\\
D_C &= g'(C)\left(\frac{\exp(\frac{a^*}{s})\left((a^*-2s)\exp(\frac{a^*}{s})+a^*+2s\right)}{s^2\left(\exp(\frac{a^*}{s})-1\right)^3}\right)\\
&\quad -g(C)\frac{\partial a^*}{\partial C}\left(\frac{\exp(\frac{a^*}{s})\left((a^*-3s)\exp(\frac{2a^*}{s})+4a^*\exp(\frac{a^*}{s})+a^*+3s\right)}{s^3\left(\exp(\frac{a^*}{s})-1\right)^4}\right)\\
D_I &= g(C)\frac{ds}{dI}\frac{\exp(\frac{a^*}{s})\left((2s^2-4a^*s+a^{*2})\exp(\frac{2a^*}{s})+(4a^{*2}-4s^2)\exp(\frac{a^*}{s})+2s^2+4a^*s+a^{*2}\right)}{s^4\left(\exp(\frac{a^*}{s})-1\right)^4}
\end{aligned}
$$

Given that e^* is a maximum, we know that $D_{e^*} < 0$, since D_{e^*} is identical to the second derivative of the TTL's maximization problem. Given the sign of D_{e^*} and conditions 4 and 5, we have $\text{sign}\left(\frac{\partial e^*}{\partial C}\right) = \text{sign}(D_C)$ and $\text{sign}\left(\frac{\partial e^*}{\partial I}\right) = \text{sign}(D_I)$, and these functions turn out to have identical positivity conditions. We find that $D_C > 0$ *and* $D_I > 0$ if and only if:

$$
\exp\left(\frac{a^*}{s}\right) > -\frac{(a^*\sqrt{4s^2+3a^{*2}}-2s^2+2a^{*2})}{(2s^2-4a^*s+a^{*2})} \tag{6}
$$

Usefully, the signs of D_C and D_I change precisely when the ratio $\left|\frac{a^*}{s}\right| = 2.3469413... \approx 2.347$, determined by plotting both functions over a^*-s space.[9] This property described by condition 6 suggests simpler rules for identifying the signs of $\frac{\partial e^*}{\partial C}$ and $\frac{\partial e^*}{\partial I}$. Substituting $a^* = g(C)(e^*-m)$ and $s = \frac{\sqrt{3h(I)}}{\pi}$ and solving for C and I, we find $\frac{\partial e^*}{\partial C} > 0$ and $\frac{\partial e^*}{\partial I} > 0$ if and only if:

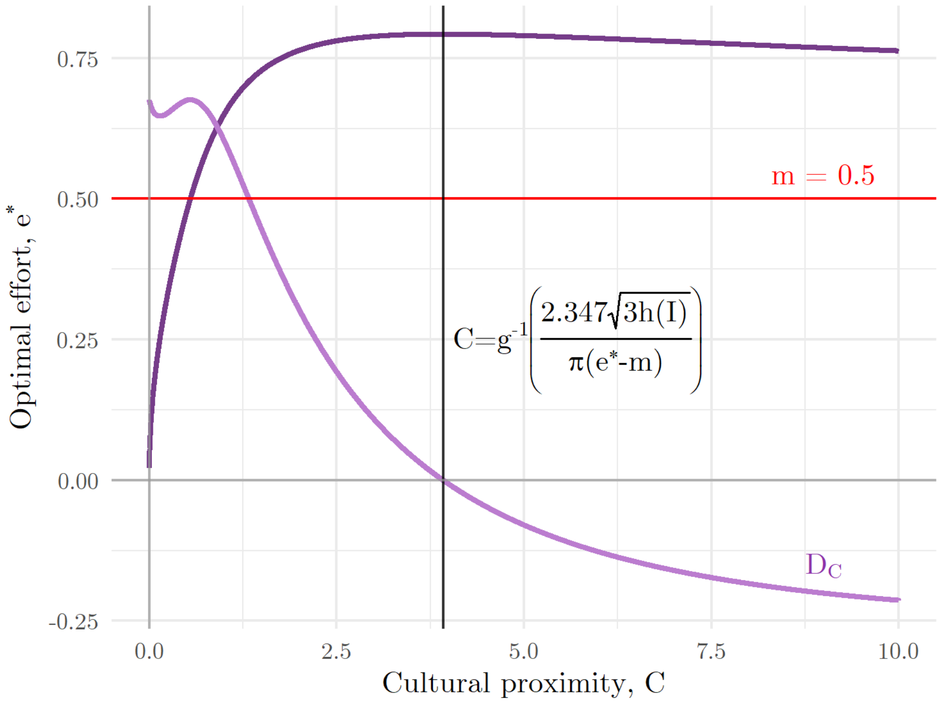
$$
C < g^{-1}\left(\frac{2.347\sqrt{3h(I)}}{\pi|e^*-m|}\right) \tag{7}
$$

$$
I < h^{-1}\left(\frac{1}{3}\left(\frac{\pi g(C)(e^*-m)}{2.347}\right)^2\right) \tag{8}
$$

where inequality 8 simply re-expresses 7. These conditions demonstrate the following predictions of the model:

[9] This is also the value at which the Einstein function $E1 = \dfrac{x^2 e^x}{(1-e^x)^2}$ achieves its positive inflection point.

Figure 1: Optimal TTL effort as a function of cultural proximity



The dark purple line plots e^* as a function of C. The lighter purple line plots D_C. The vertical line marks the point where $\frac{\partial e^*}{\partial C}$ becomes negative. $I = 5$, $m = .5$, $g(C) = \sqrt{C}$, $h(I) = I^{-1}$.

1. Increased cultural proximity or increased institutional quality make TTL effort more observable to the principal.

2. Increased observability motivates effort, but only up to a point. After that point, increased observability leads to a decrease in effort.

These predictions capture intuitive notions about how humans choose to exert effort. When our effort is not very likely to be recognized and rewarded, we will put in little effort. If our effort is more observable, but we might be blamed for factors beyond our control—i.e., we might receive an unfavorable η—we "hedge" against bad fortune by working harder than we are asked to work. But as our effort becomes still more effective, it is no longer necessary to work quite so hard. This story plays out in figure 1, which plots the TTL's optimal effort over various levels of cultural proximity, holding m and I constant.

4 Empirical strategy

With satisfactory measures of cultural proximity between the TTL and the recipient country, the institutional quality of the recipient country, and the success of the project outcome, the predictions of the model can be tested empirically.[10] The basic estimating specification will be:

$$\begin{aligned} \mathit{Outcome}_i = \beta_0 + \beta_1 \mathit{Institutional\ quality}_i + \beta_2 \mathit{Cultural\ proximity}_i + \\ \beta_3(\mathit{Institutional\ quality}_i \times \mathit{Cultural\ proximity}_i) + X_i'\gamma + \varepsilon_i, \end{aligned} \tag{9}$$

where $\mathit{Outcome}_i$ is an ordinal measure of success for project i, $\mathit{Institutional\ quality}_i$ is a measure of the institutional quality in project i's recipient country in the year the project was approved, $\mathit{Cultural\ proximity}_i$ is the negative of genetic distance between the TTL assigned to project i and the recipient country for project i, rescaled such that standard deviation of $\mathit{Cultural\ proximity}_i = 1$. Finally, X_i is a vector of other controls. I choose a linear specification to allow for a rich set of fixed effects and an instrumental variables strategy. The construction of these variables of interest, and the variables included as additional controls, are described in more detail in the sections below.

4.1 Cultural proximity

I use a two-step process to estimate a TTL's cultural proximity relative to the recipient country. First, I use TTL surname to determine their likely country of ancestry. Then, I need a measure of genetic distance between the TTL's home country and the project's recipient country to use as a proxy for (negative) cultural proximity. The measure of genetic distance I use was developed by Pemberton et al. (2013) to describe differences in microsatellite variation between human populations, and then aggregated by Wacziarg and Spolaore (2018) to describe genetic distance between entire countries.[11] The raw data assembled by Pemberton et al. (2013) provide pairwise genetic distances between 267 genetic populations.

[10] This section presents only the reduced form empirical strategy. Appendix G describes my first steps toward structural estimation. Eventually sections 4 and 5 will be organized primarily around the structural methods and results.

[11] Microsatellites are sequences of repeating base pairs in DNA. Microsatellites have a higher mutation rate that other areas of DNA, making them ideal for measuring genetic diversity between populations.

Wacziarg and Spolaore (2018) match those populations to ethnic groups in each country to aggregate these genetic distances to the country level, using country-level ethnic composition data from Alesina et al. (2003). The measure of genetic distance used by Pemberton et al. and adapted by Wacziarg and Spolaore is based on genetic differences that geneticists believe to be neutral, in the sense that these differences are uncorrelated with genetic fitness and are not affected by genetic selection pressures.[12]

As further background, the Alesina et al. (2003) data set includes population proportions of 1120 distinct ethnic groups in each country. Wacziarg and Spolaore match each ethnic group to one of the 267 genetic populations in Pemberton et al. (2013)'s genetic distance data to calculate pairwise, country-level genetic distance. With minor adjustments, I rely on the Wacziarg and Spolaore data set as the basis for the subsequent calculations of my new cultural proximity measure in the context of World Bank TTLs and their assigned recipient countries.

To construct my new variable, I need to know each TTL's country of origin to determine their cultural proximity to the population of the relevant recipient country. Unfortunately, this information is not available in any public archive, although the name of each TTL is a matter of public record. I use TTL surnames to construct probability-weighted collections of likely ancestral countries. I first gather data on the global distribution of each TTL surname from *Forebears*, a globally representative surname database.[13] Most surnames are found in more than one country. Where this is the case, I construct my cultural proximity measure as the negative of the weighted average of the genetic distances between recipient country and the 10 most likely ancestral countries based on TTL surname, where the weights are the proportion of all individuals with the TTL's surname who live in each country. I assume here that each TTL's ancestral country represents a random draw from the global distribution of

[12]In previous work Wacziarg and Spolaore (Spolaore and Wacziarg, 2009) use a measure of genetic distance aggregated to the country level from 42 genetic populations. Their data based on the work of Pemberton et al. (2013) represents a significant improvement in the granularity of their resulting data. Two countries, Togo and Tanzania, are not included in Wacziarg and Spolaore's updated data set for 2018. I regress Wacziarg and Spolaore's 2009 measure of genetic distance on their new measure and use interpolated values of the new measure for these countries.

[13]*Forebears* is a fairly new resource, having been launched in 2012, but already there is some precedent in academic research for using it, as I do, to determine likely country of origin for individuals based on surname. See Nguyen et al., 2017; Pursiainen, 2019.

individuals with the same surname. If this assumption is approximately met, my constructed measure of cultural proximity introduces only attenuation bias in my OLS and IV coefficient estimates, relative to the information I could use if data were available concerning each TTL's actual history of residence. Note that the maximum value of *cultural proximity* is 0 because it is constructed as the negative of *genetic distance*, an intrinsically non-negative measure. A cultural proximity of 0 would indicate that a TTL has a surname found exclusively in their project's recipient country.[14]

4.2 Institutional quality

As a measure of institutional quality within each country, I use the Country Policy and Institutional Assessment (CPIA) provided by the World Bank. For each project, I furthermore use the CPIA of the recipient country *in the year* the project was approved. The CPIA, measured on a six-point ordinal scale, is the simple average of four six-point "cluster" scores along the separate dimensions of economic policies, structural policies, quality of public administration, and social inclusion.[15]

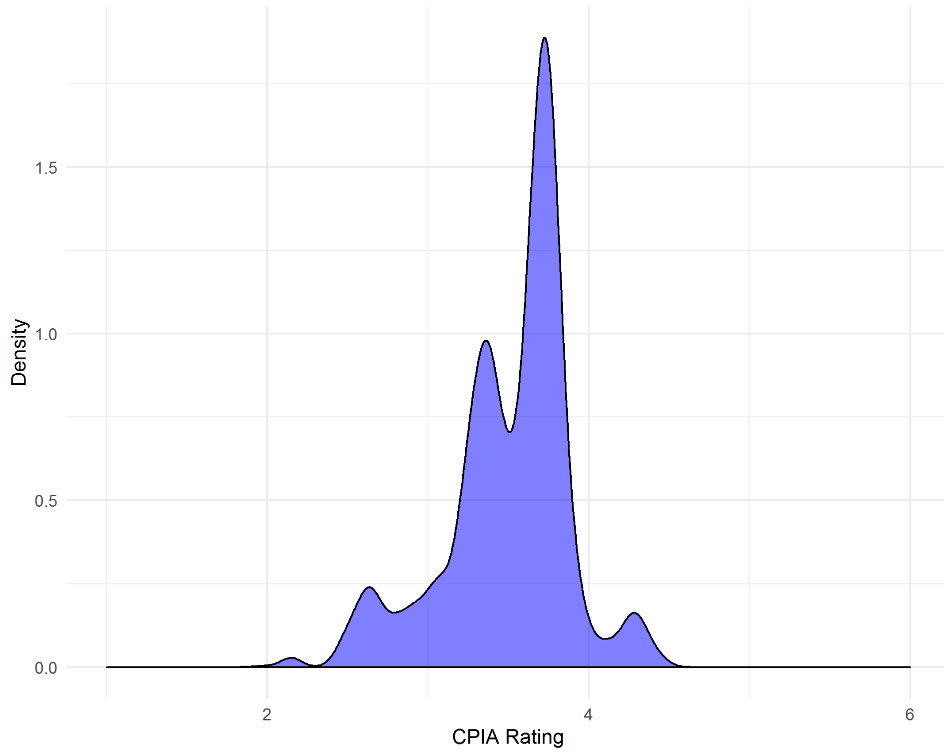
The World Bank began measuring CPIA in the mid-1970s, and currently determines a CPIA value for nearly every country. Unfortunately, CPIA ratings before 2005 are not publicly available. Further, CPIA ratings are only provided for the 95 economies the World Bank classifies as being "low income." Given that the CPIA measure is central to my analysis, I drop all projects for which no CPIA data are available, essentially excluding all middle-income countries, despite the fact that several of these countries (e.g. Argentina, Brazil, China, Poland, South Africa) qualified for World Bank assistance at some point after 2005.

For projects that begin before 2005, but were evaluated for their success after 2005, I use the average CPIA rating for the country in question over all years for which data are available. Imputing CPIA data in this way is not ideal, but country-level CPIA ratings are relatively stable over time, so the average of the CPIA in available years is likely to be a

[14] My use of surnames to infer cultural proximity may systematically over- or under-estimate cultural proximity for female TTLs because it is customary in many countries for married women to adopt their husband's surname. I address this concern in Section 6.

[15] I also consider specifications that employ each of the six-point cluster scores individually in place of the CPIA average. These results do not differ qualitatively and are reported in Appendix E.

Figure 2: Distribution of sample projects' CPIA ratings



reasonable approximation when data are missing.[16]

4.3 Outcome variable: project success

After a World Bank foreign aid project is completed or abandoned, the IEG evaluates the project on several dimensions and determines an overall outcome score, which is, like CPIA, measured on a six-point scale. I use this overall ordinal outcome score as my main dependent variable.

[16]For the 75 countries represented in my sample, CPIA data are on average available for 11.95 years (among 13 possible years, 2005-2017). The average standard deviation of CPIA within each country is 0.12 for the available years. Zimbabwe's CPIA fluctuates the most, with a standard deviation of 0.47. Excluding Zimbabwe from my sample and repeating my analysis yields qualitatively similar results.

Though IEG evaluators strive to be objective in their assessments, some subjective judgments are unavoidable in project evaluation. Each project is evaluated against its own stated objectives, and the objectives are themselves evaluated for their relevance to the needs of the recipient country. While a universal and objective measure of project success would be preferable, no such measure exists, so IEG evaluations have become a widely used proxy for project success (see Geli et al., 2014, for a concise review).[17]

The IEG provides data on every project it has ever evaluated, including outcome score, location, cost, year and project ID. This last variable allows me to link projects uniquely with other data provided by the World Bank, including the names of the TTLs.

The IEG updated its evaluation methodology in 2005, so I limit my sample of projects to those evaluated in 2005 or later. I also drop projects that were rated "Not Applicable" or "Not Rated," or for which the necessary data were otherwise incomplete.

The IEG rates each project at one of six levels, which I convert to a numerical score: *highly unsatisfactory* (1), *unsatisfactory* (2), *moderately unsatisfactory* (3), *moderately satisfactory* (4), *satisfactory* (5), and *highly satisfactory* (6). Nearly half of all projects in my sample are rated as *moderately satisfactory.* The name of the evaluator is given in the IEG's published assessment, called the Implementation Completion and Results Report Review (ICRR). This information allows me to control for evaluator fixed effects.[18]

4.4 Other controls

I also include a number of control variables in my estimation specification. For example, it is possible that some global regions simply produce better TTLs than others, or that some regions are more likely to host successful aid projects due to factors not accounted for by other variables. To control for this, I employ fixed effects for the project region and the TTL home region. Project region is designated in each IEG evaluation, and TTL home region is, like my measure of cultural proximity, a weighted measure based on the global distribution of individuals who share a TTL's surname.

[17] In defending their decision to use IEG evaluation data, Chauvet, Collier and Duponchel (2010) write, "the evaluation procedure is independent, staff are experienced, the process has been on-going for more than three decades, and a lot of resources are put into it."

[18] See Appendix B for details about how the IEG assigns ratings.

Table 1: Summary statistics for high- and low-cultural proximity projects

	High proximity		Low proximity		Full sample	
	Mean	*St Dev*	*Mean*	*St Dev*	*Mean*	*St Dev*
Cultural proximity	-0.57	0.42	-2.33	0.51	-1.45	1
CPIA	3.54	0.39	3.47	0.38	3.50	0.39
IEG outcome	4.00	0.96	3.79	0.99	3.89	0.98
Approval year	2004.57	4.18	2004.85	4.38	2004.71	4.28
Evaluation year	2012.16	3.63	2011.94	3.67	2012.05	3.65
N	973		973		1946	

Notes: Summary statistics for projects divided into subsamples at median cultural proximity.

TTL experience may also be relevant. To capture the effect of TTL experience on project success, I include an indicator variable equal to 1 if that TTL has completed at least one other project prior to the approval date of the current project, even if that project does not appear in my sample.

I control for the cost of each project, a variable that is included in the IEG's data set. Very expensive projects may be more complicated and therefore prone to failure. Alternatively, large projects, when they succeed, may have positive effects that are especially salient to IEG reviewers, leading to higher outcome scores. In either case, cost is likely to be an important control. The marginal effect of project cost on project outcome is likely to decline as cost increases, so I include the logarithm of project cost in each regression.

Similarly, I control for the recipient country's population in logs. Countries with larger populations may pose larger challenges for foreign aid projects, but the marginal effect of population size is likely to decrease as population increases.

I calculate growth in each country's GDP per capita (in terms of purchasing power parity, PPP) over the lifetime of each project. Rather than use the simple average of the yearly growth rates, I use the GDP per capita (PPP) in the first and last year of the project—as given by the World Bank—and the duration of the project in years to calculate the counterfactual growth rate as if this rate had been constant over the life of the project. This procedure yields growth estimates that are very similar to the simple average, but "penalizes" countries that experience uneven growth.[19] I include this control variable because heterogeneous growth rates across countries are a potential source of variation in project outcomes that is partly unpredictable by the TTL or the World Bank.

Following Kaufmann, Kraay and Denizer (2013), I include in nearly every regression a set of indicator variables that captures the time interval when the project began, the time interval when it was evaluated, the project's sector, and interactions between sector and the two timing variables. Specifically, I control for the project's approval date, in five-year bins, to capture changes over time in the mix of projects the World Bank approves. I also control for project evaluation date, likewise in five-year bins, as standards for evaluation may have shifted over time. Additionally, because both kinds of changes may vary by sector, I interact both groups of year bins with a set of indicator variables designating project sector. A project's sectors are specified in its IEG evaluation. I classify each project as belonging to the sector listed first in the IEG evaluation.[20]

I also control for the length of time between a project's conclusion and its evaluation. Most projects are evaluated between two and three years after completion, but some are evaluated much later, which may limit the information available to evaluators and add noise to the assigned outcome score. Alternatively, a longer delay between project conclusion and evaluation may bring the project's longer-term effects into clearer focus. In either case, it may be important to control for any evaluation lag.

Finally, for the IEG rating of project success that serves as my outcome variable, two kinds of evaluations appear in my sample: ICRRs and PARs.[21] Both evaluation types use the same criteria and are intended to be broadly comparable. However, PARs are based on a more thorough review of evidence related to project success, and are issued by a group rather than by a single IEG reviewer. PARs also tend to assign slightly lower outcome scores, so I

[19] For instance, if a country's GDP is halved one year, and then doubled the next year, the simple average of its GDP growth rates over the two years is 25% (the mean of -50% in the first year and 100% in the second). The counterfactual-constant growth rate over these two years is 0%, as the starting and ending annual GDPs are identical.

[20] The project sectors that appear in my data set are Agriculture and Rural Development; Economic Policy; Education; Energy and Mining; Environment; Finance; Global Information/Communications Technology; Health, Nutrition and Population; Public Sector Governance; Social Protection; Transport; and Urban Development. Additionally, 29 project sectors are listed as "Unknown."

[21] See Appendix B for descriptions of these two evaluation types.

include an indicator variable that distinguishes between these two evaluation types.

4.5 Potential endogeneity of TTL cultural proximity

Estimation results from ordinary least squares regression cannot be interpreted as causal effects. The biggest threat to causal identification is that assignment of TTLs to projects may be endogenous. For example, if culturally close TTLs are preferentially assigned to more difficult projects, OLS will bias toward 0 the coefficients on $\textit{Cultural proximity}_i$ and its interaction with $\textit{Institutional quality}_i$. Alternatively, culturally proximate TTLs may be assigned to particularly difficult projects in countries with low institutional quality, but they may be assigned to the easiest projects in countries with high institutional quality. Despite the theorized prediction, such an assignment pattern could conceivably account for the observed positive coefficients on the interaction term in models 2 and 3 in Table 2, even if the true causal effect of the interaction term on project quality were 0.

To address this concern, I use an instrumental variables strategy to isolate the variation in cultural proximity explained by conditions exogenous to the characteristics of the recipient country: the availability of culturally close TTLs. As an instrument, I use the average cultural proximity to that project's recipient country of all TTLs who led other projects that are both (a) located in low-income countries and (b) evaluated *before* 2005. That is, I use the average cultural proximity of TTLs who led projects that are excluded from my estimating sample only because they were evaluated before the starting date for my sample. I refer to this set of projects as "pre-sample" projects. By construction, this instrument varies between countries but not within countries. The instrumental variable's value is identical for every project in a given country, so the results in Table 2 cannot be an artifact of a TTL assignment policy that simply gives the most difficult projects to culturally proximate TTLs in countries with low institutional quality and easier projects to culturally proximate TTLs in countries with high institutional quality.

Conceptually, the average cultural proximity of the earlier pre-sample TTLs to the recipient country for a given aid project should provide a good summary statistic for the availability of culturally close TTLs for projects in that country. Where culturally close TTLs are scarce, such that average cultural proximity of the earlier TTLs is low, we should expect, on average, to see a more culturally *distant* TTL assigned. The first-stage results (presented in Table 3) support this claim. To overcome endogeneity, however, instrumental variables (IVs) need to meet the usual exclusion restriction—they need to be uncorrelated with the main regression's error term. In practice, this means the IV should be related to the outcome variable only through the endogenous variable (or through other included exogenous regressors). My IV varies only between countries (every project in a given recipient country has the same average cultural proximity to the fixed set of pre-sample TTLs). Thus my instrument is guaranteed to eliminate any source of endogeneity arising from within-country variation, and we need be concerned only with between-country sources of endogeneity in my IV models.

There are two reasons why the instrument may fail the exclusion restriction because of between-country sources of endogeneity. I examine the plausibility of each concern.

First, an abundant supply of culturally close TTLs (relative to a given country) may result from the World Bank's unequal involvement with various countries or cultures. The World Bank may seek out or cultivate talented individuals from favored countries or cultures. If projects in those preferred countries or cultures also receive other support that promotes project success, the IV fails the exclusion restriction. However, we have two good measures of the World Bank's investment in individual countries—the number of projects they complete in each country and the amount they spend in each country. To determine whether the supply of TTLs that are culturally close to a country is correlated with observable measures of World Bank investment in that country, I regress the instrument on the number of projects in each recipient country since 1990, controlling for the logarithms of population and GDP per capita, both of which I include as controls in every regression. I find that this cumulative number of projects does not predict average TTL cultural distance (p-value=.89). A similar reduced-form regression, including the total cost of all projects in each recipient country since 1990, also yields an insignificant effect of total project cost on the instrument (p-value=.71). The results for both of these regressions are reported in Appendix C, Table C1. These results support the contention that if the World Bank favors some countries over others, it does not express this favoritism through differential recruitment of people to serve as TTLs, so the exclusion restriction appears not to be violated in this manner.

A second concern is that cultures that are most conducive to project success may also produce a disproportionate number of qualified World Bank applicants. A culture's strong emphasis on the value of education, for example, may lead to a large pool of qualified, motivated World Bank applicants from that culture. If those same cultural traits also produce conditions which are conducive to project success, the proposed IV may fail the exclusion restriction. Fortunately, this concern also appears to be unfounded. The best measure available for conditions that may be conducive to project success is the CPIA measure, which is intended to capture explicitly the suitability of a country's institutions for development projects and poverty-reduction projects. Institutions are not the same as culture, but where culture facilitates World Bank project success, this likely occurs at least partly through the culture's influence on institutions. Again, regressing the proposed instrument on each recipient country's average CPIA, controlling for the logarithms of population and GDP per capita, I find no statistically significant relationship between CPIA and this instrument (p-value=.63). These results are also included in Table C1 of Appendix C.

5 Results

I begin by estimating the effects of cultural proximity and institutional quality on project success using OLS methods. After presenting the basic OLS results, I estimate an instrumental variables specification.

According to the conceptual model described in Section 3, cultural proximity should be beneficial for project outcomes unconditional on institutional quality, implying $\beta_2 > 0$ in the empirical model in equation (11). If we force $\beta_3 = 0$ by excluding the interaction term from the equation (11), as in model 1 in Table 2, we indeed find that $\beta_2 > 0$, i.e. that cultural proximity is associated with improved project outcomes, independent of institutional quality. In Table 2, model 2, we allow the effect of cultural proximity on project success to vary with institutional quality by including the appropriate interaction term in the regression. We find a significant and positive coefficient on the interaction term ($\beta_3 = 0.062$, $p < .01$). This coefficient implies that cultural proximity is *more beneficial* in countries with strong institutions than in those with weak institutions.

The effect sizes of cultural proximity and its interaction with institutional quality are moderate. At the mean institutional quality, a one-standard-deviation increase in cultural proximity implies a 0.067 increase in IEG project score on a six-point scale. This is about two-thirds as large as the effect size of institutional quality. This seems economically important, given that institutional quality has been repeatedly identified in the literature as a strong predictor of the success of foreign aid projects.

Table 2: OLS results, selected coefficients

	Dependent variable: Project success		
	(1)	(2)	(3)
Cultural proximity	0.069**	0.068**	0.063*
	(0.033)	(0.033)	(0.036)
Institutional quality	0.096***	0.192***	0.198***
	(0.030)	(0.049)	(0.050)
Cultural proximity ×		0.062**	0.055**
Institutional quality		(0.026)	(0.027)
Other controls	Yes	Yes	Yes
Reviewer FEs	No	No	Yes
N	1,946	1,946	1,946

Notes: *** < .01, ** < .05, * < .1

Project success is IEG's evaluation of the project on a 1-6 scale. Institutional quality is CPIA standardized to mean = 0, sd = 1. Cultural proximity is re-scaled such that sd = 1. Standard errors are clustered by country-evaluation year.

Each model includes controls for project cost, TTL experience, project region, probability-weighted average of TTL's likely ancestral region, recipient country GDP per capita (log) and GDP growth, recipient country population (log), project sector (as defined by the World Bank), project approval year, project evaluation year, and the length of time between the end of the project and its evaluation.

5.1 IV results

Table 3 presents first-stage results for the main IV specification. My analysis relies on cultural proximity and its interaction with institutional quality, so I need two instruments and two first-stage equations, one to explain *cultural proximity* and one to explain *cultural proximity* × *institutional quality*. My second instrument, then, is just my first instrument (*Pre-sample average TTL cultural proximity*) interacted with *Institutional quality*. Other than the two instruments, all first stages include the same exogenous controls as their second stage counterparts.

The F statistics for the joint significance of the pair of instruments in the first-stage regressions are 93.9 ($p < .001$) and 221.8 ($p < .001$) in models (1) and (2), respectively, indicating that the instruments are highly relevant. Furthermore, the instruments function in the manner expected: *pre-sample TTL cultural proximity* is an excellent proxy for the actual *cultural proximity* between TTL and recipient country and *pre-sample TTL cultural proximity* × *institutional quality* is highly predictive of *cultural proximity* × *institutional quality*, and both of the key coefficients are positive as expected.

Table 4 presents results for the second-stage instrumental variable regressions. Model 2 gives the IV results for my main specification. The estimated effect of the interaction term is almost twice the magnitude estimated by OLS (model 3 in Table 2). The estimated coefficient on *cultural proximity*, representing the marginal effect of cultural proximity at the mean institutional quality (since CPIA is standardized), is also much larger, but is significant only at the 10% level. Why might the IV regression produce a larger estimate for the effect of cultural proximity? If the World Bank wants to ensure that projects meet a minimum level of success and operates under the impression that culturally close TTLs have an advantage in producing high-quality projects, then assigning culturally close leaders to difficult projects, no matter the recipient country's institutional quality, would be a sensible policy. This assignment strategy would bias downwards—toward zero—the OLS coefficient on *cultural proximity*, since culturally proximate TTLs would be assigned to more difficult projects. If some components of project difficulty are unobservable to the econometrician (but observable to World Bank staff), and therefore cannot be controlled for, then the positive effect of *cultural proximity* will be partially offset by endogenous TTL assignment. As a result, the OLS estimate for *cultural proximity* × *institutional quality* will also be biased

Table 3: IV: First stage regressions (selected coefficients)

	Cultural proximity	Cultural proximity $\times$ Institutional quality
	(1)	(2)
Pre-sample TTL cultural proximity	0.455^{***}	-0.028
	(0.034)	(0.045)
Institutional quality	-0.083	0.267^{***}
	(0.053)	(0.094)
Pre-sample TTL cultural proximity $\times$ Institutional quality	-0.041^{***}	0.550^{***}
	(0.015)	(0.030)
Other controls	Yes	Yes
Observations	1,946	1,946
R^2	0.597	0.838
Joint F-stat for IVs	93.9^{***}	221.8^{***}

Notes: $^{***} < .01$, $^{**} < .05$, $^{*} < .1$

For a summary of additional control variables included in all models, see notes to Table 2.

toward zero. The increased magnitudes of the IV coefficient estimates, relative to the OLS estimates, are not surprising because the IV estimates should better reflect the true causal effects of *cultural proximity* and *cultural proximity* × *institutional quality* in the absence of an endogenous assignment strategy that would be expected to attenuate OLS estimates.

Model 3 includes fixed effects for the 230 IEG reviewers who rated the success of projects in my sample. As noted above, projects are rated according to a standardized set of criteria, but there remains an element of subjectivity in IEG outcome scores. Some reviewers may be "tougher" than others, and if toughness is correlated with institutional quality or cultural proximity, estimates will be biased. Model 3 provides evidence which suggests that this concern is unfounded.[22]

Table 5 demonstrates that the results do not change qualitatively if I use an alternative measures of institutional quality, the World Governance Indicators (WGI) (Kaufmann et al., 2011). The WGI is a set of six separate measures of institutional quality: voice and accountability, political stability and absence of violence, government effectiveness, regulatory quality, rule of law, and control of corruption. Crucially, the WGI is not specifically designed to measure the suitability of a country for aid projects, and the World Bank does not use the WGI to allocate resources. Table 5 is analogous to Table 4, but uses the average of the six WGI indicators (standardized to $\mu = 0$, $\sigma = 1$) in place of the CPIA as the measure of institutional quality. Using WGI in place of CPIA yields comparable results for the interaction term *Cultural proximity* × *Institutional quality*, but the effect of *Cultural proximity* at the mean WGI is not statistically significant, even at the 10% level.

6 The possibility of mismeasurement for female TTLs

Given that my measure of cultural proximity is based on TTL surnames, cultural proximity for female TTLs may be systematically mismeasured. Specifically, female TTLs may take a new surname at marriage, and their new surname may not be representative of their

[22]If a project's rating comes from a more rigorous Project Performance Assessment Report (PAR), which is conducted by a group, I categorize the reviewer as $PAR = 1$ so that all such projects are given the same fixed effect. Fifteen observations had no reviewer listed, so I assign these observations the same fixed effect. Additionally, 72 observations were evaluated by reviewers who reviewed no other projects in my sample. These 72 observations are assigned a single fixed effect.

Table 4: IV results, selected coefficients

	Dependent variable: Project success		
	(1)	(2)	(3)
Cultural proximity	0.185	0.205*	0.185
	(0.122)	(0.121)	(0.123)
Institutional quality	0.095***	0.261***	0.313***
	(0.030)	(0.074)	(0.078)
Cultural proximity ×		0.107**	0.130**
Institutional quality		(0.046)	(0.048)
Other controls	Yes	Yes	Yes
Reviewer FEs	No	No	Yes
N	1,946	1,946	1,946
Notes:			*** < .01, ** < .05, * < .1

For a summary of additional control variables included in all models, see notes to Table 2.

Table 5: IV results, using WGI in place of CPIA

	IEG Outcome: 1-6 scale		
	(1)	(2)	(3)
Cultural proximity	0.183	0.146	0.117
	(0.124)	(0.123)	(0.125)
Inst. quality (WGI)	0.095***	0.259***	0.301***
	(0.026)	(0.064)	(0.067)
Cultural proximity ×		0.108***	0.130***
Inst. quality (WGI)		(0.040)	(0.042)
Other controls	Yes	Yes	Yes
Reviewer FEs	No	No	Yes
N	1,946	1,946	1,946

Notes: *** < .01, ** < .05, * < .1

WGI is standardized to $\mu = 0$, $\sigma = 1$. For a summary of additional control variables included in all models, see notes to Table 2.

cultural background. In this section I provide evidence, using a second measure of cultural proximity based on TTLs' given names, that this possibility is unlikely to influence my findings significantly and report the key coefficients of my main IV specification with controls for gender.

Table 6: Correlation between measures of cultural proximity based on given name and surname

	Full sample	Men only	Women only
Correlation	0.2630	0.2647	0.2606
95% CI	(0.2165, 0.3084)	(0.20756, 0.3201)	(0.1792, 0.3385)
N	1578	1051	527

Notes: Correlation is Pearson's product-moment correlation. The total number of observations does not equal 1946 because instances in which a TTL oversees multiple aid projects in the same country yield duplicate observations, which are removed. Additionally, observations are removed for TTLs who use only a first initial because a first initial is likely to be wholly uninformative about cultural background.

To test empirically the possibility that female TTLs' surnames are less informative than those of male TTLs, I reconstruct my cultural proximity measure using *given names* instead of surnames and compare this new variable to my original measure of cultural proximity. This new measure of cultural proximity is constructed in precisely the same way as the original measure, but I draw the global distribution of each TTL's given name from Forebears.io to determine a set of probable home countries for each TTL.[23] We should expect these two measures of cultural proximity to be correlated for both male and female TTLs. If the correlation between these two measures of cultural proximity is significantly higher for male TTLs than for female TTLs, then we have evidence that female TTLs' surnames may yield a mismeasurement of cultural proximity. Specifically, this analysis relies on the following assumptions:

1. Given names are somewhat informative about cultural background.

[23]See section 4.1 for details on the construction of my main cultural proximity measure.

2. Given names are equally as informative in this respect for women as for men.

3. Individuals—men or women—do not change their given names upon marriage.

TTLs' genders are unknown, so I use their given names to infer gender. I processed each given name using Genderize.io, a database of the gender rates for more than 200,000 given names, which predicted gender for most TTLs.[24]

Some given names could not be matched by the Genderize.io database. For these remaining 293 TTLs who have unusual given names, gender-neutral names, or only a first initial, I use publicly available online sources to determine the gender of that particular World Bank employee. For each TTL gender determined in this way, I recorded the URL of the source.

As reported in Table **??**, the correlation between cultural proximity calculated based on each TTL's given name and cultural proximity calculated based on each TTL's surname is positive and significantly different from 0 for both the male and female subsamples. Furthermore, these correlations are not significantly different from each other, suggesting that cultural proximity is not mismeasured for female TTLs relative to male TTLs.[25]

7 Conclusion

This study examines the relationships between the cultural proximity between World Bank TTLs and foreign aid recipient countries, the institutional quality of recipient countries, and the success of foreign aid projects. Examining these relationships poses two problems. First, there is no existing measure of cultural proximity between TTLs and recipient countries. Second, the assignment of TTLs to recipient countries is likely to be endogenous. I overcome the first of these challenges by constructing a plausible measure of cultural proximity from an existing measure of genetic distance between countries and the global distribution of surnames. I overcome the second of these challenges by instrumenting for cultural proximity with a measure of the availability of culturally close TTLs.

[24]See Topaz and Sen (2016), Greenberg and Mollick (2015), and Mohammadi and Shafi (2018) for other instances of research using Genderize.io to infer subject gender.

[25]Appendix D reports coefficient estimates from my main IV specification with an indicator variable for female TTLs interacted with the coefficients of interest.

I find evidence of a positive causal relationship between the cultural proximity of foreign aid project leaders and their projects' outcomes. Furthermore, this relationship appears to be significantly stronger in countries with strong institutions. These findings imply that multilateral aid organizations should give special attention to the cultural background of project leaders they assign in countries with high institutional capacities.

A Data appendix

A.1 Genetic distance

Wacziarg and Spolaore calculate genetic distance between countries according to the following formula:

$$F_{st}^{12} \sum_{i=1}^{I} \sum_{j=1}^{J} (s_{1i} \times s_{2j} \times d_{ij}),$$

where $i = 1, ..., I$ indexes the genetic groups in country 1, $j = 1, ..., J$ indexes those of country 2, s_{1i} and s_{2j} are the population shares of genetic groups i and j in countries 1 and 2 as given by Alesina et al. (2003), and d_{ij} is the genetic distance between i and j as given by Pemberton et al. (2013). F_{st}^{12} then is the expected genetic distance between two randomly selected citizens, one from each country.

A.2 IEG evaluation

Every World Bank aid project receives at least two ratings. First, the TTL evaluates their own project according to the same criteria the IEG uses. This preliminary evaluation, called an Implementation, Completion and Results Report (ICR), and the documents on which it is based, are then reviewed by an IEG evaluator, who issues a second report—an ICR Review (ICRR). The ICRR may revise the outcome score given in the initial ICR. Where insufficient supporting documentation is provided, IEG evaluators are instructed to penalize the project's assessment.

Some projects receive a third evaluation—a Project Performance Assessment Report (PAR). These evaluations follow the same rating criteria as ICRRs, but are based on a broad range of independently gathered evidence, including site visits and interviews with stakeholders. I use the PAR score if it is available, and otherwise use the ICRR score. ICR scores are determined by the TTL in charge of the project, so I do not use ICR scores in my analysis. Because PARs are far more rigorous than ICRRs and represent the research and judgments of a group of evaluators, I assign every PAR the same fixed effect.

B Structural estimation results (preliminary)

To test the validity of my principal-agent model, outlined in section 3, I use the model's production function as the estimating equation:

$$Q = q(X) + g(C)e + \eta$$

where $q(X)$ is a linear combination of control variables, and e is predicted utility maximizing level of effort for each TTL, given their cultural proximity to their project's recipient country and the institutional quality of the recipient country. Additionally, a number of other parameters that govern the TTL's choice of effort require estimation. I choose the following simple functional forms for $g(C)$ and $h(I)$:

$$g(C) = \gamma_1 C^{\gamma_2}$$
$$h(I) = \lambda_1 + \lambda_2 I$$

where γ_1, γ_2, λ_1, and λ_2 are parameters to be estimated, with the restrictions that $\gamma_1 > 0$, $0 < \gamma_2 < 1$, $\lambda_1 > 0$, $\lambda_2 < 0$, and $\lambda_1 > \lambda_2 \bar{I}$, where $\bar{I}$ is the largest observed value of I. The threshold effort level m is also a parameter to be estimated.

To limit the parameter-space to a manageable dimensionality, I impose the following function for the TTL's disutility of effort:

$$\phi(e) = \frac{1}{2}e^2$$

I re-scale C and I by subtracting each variable's (negative) minimum observed value and then adding 0.1, so that $C > 0$ and $I > 0$. This re-scaling preserves the feature that a 1-unit increase in C or I is associated with a one-standard deviation increase.

Structural estimation proceeds iteratively as follows.

Step 1

Begin with guesses for values of the parameters in $h(I)$: λ_1 and λ_2.

Step 2

Using these values of λ_1 and λ_2, run a numeric optimization routine to find values of γ_1, γ_2, and m that minimize the residual sum of squares from OLS estimation of the production function.[26] Values of e for each project are determined within the optimization routine. For the OLS specification, $q(X)$ includes precisely the same set of variables described in section 4 and used in estimating the reduced form results, with the exception of the interaction between C and I. The quantity $g(C)e$ is included as a regressor with its coefficient restricted to 1.

Step 3

Save the residuals from OLS in the last step of OLS before convergence. Group these residuals into 50 bins by their associated value of I and calculate the variance within each bin. Then regress these 50 variances on each bin's mean value of I. The resulting OLS estimates provide new values for λ_1 and λ_2, the parameters in $h(I)$, so that the variance in the residuals (as a function of I) provides an estimate of the variance in η (as a function of I). Return to step 2.

This process terminates when two successive iterations produce estimates for λ_1, λ_2, γ_1, γ_2, and m that are sufficiently close. The stopping criterion I've been using is:

$$\sqrt{(\lambda_1^t - \lambda_1^{t-1})^2 + (\lambda_2^t - \lambda_2^{t-1})^2 + (\gamma_1^t - \gamma_1^{t-1})^2 + (\gamma_2^t - \gamma_2^{t-1})^2 + (m^t - m^{t-1})^2} < 5 \times 10^{-5}$$

where t indexes iterations.

The parameter estimates resulting from this estimation process are given in Table G1. Figure 3 plots the estimated effort levels for each project at the last iteration. Figure 4 plots the variances of the residuals within each bin at the last iteration. Comparing Akaike information criteria for the structural model (5320.09) and the reduced form model estimated in Table 2, column 3 (5325.757) shows the structural model to be an improvement.

[26] I recover similar parameter estimates using Nelder-Mead, Broyden–Fletcher–Goldfarb–Shanno, or Conjugate Gradient methods. I present results from Nelder-Mead, which converges most quickly.

Table B1: Structural parameter estimates

Parameter	Estimate
λ_1	1.278
λ_2	-0.128
γ_1	1.460
γ_2	0.417
m	0.934

Figure 3: Predicted effort

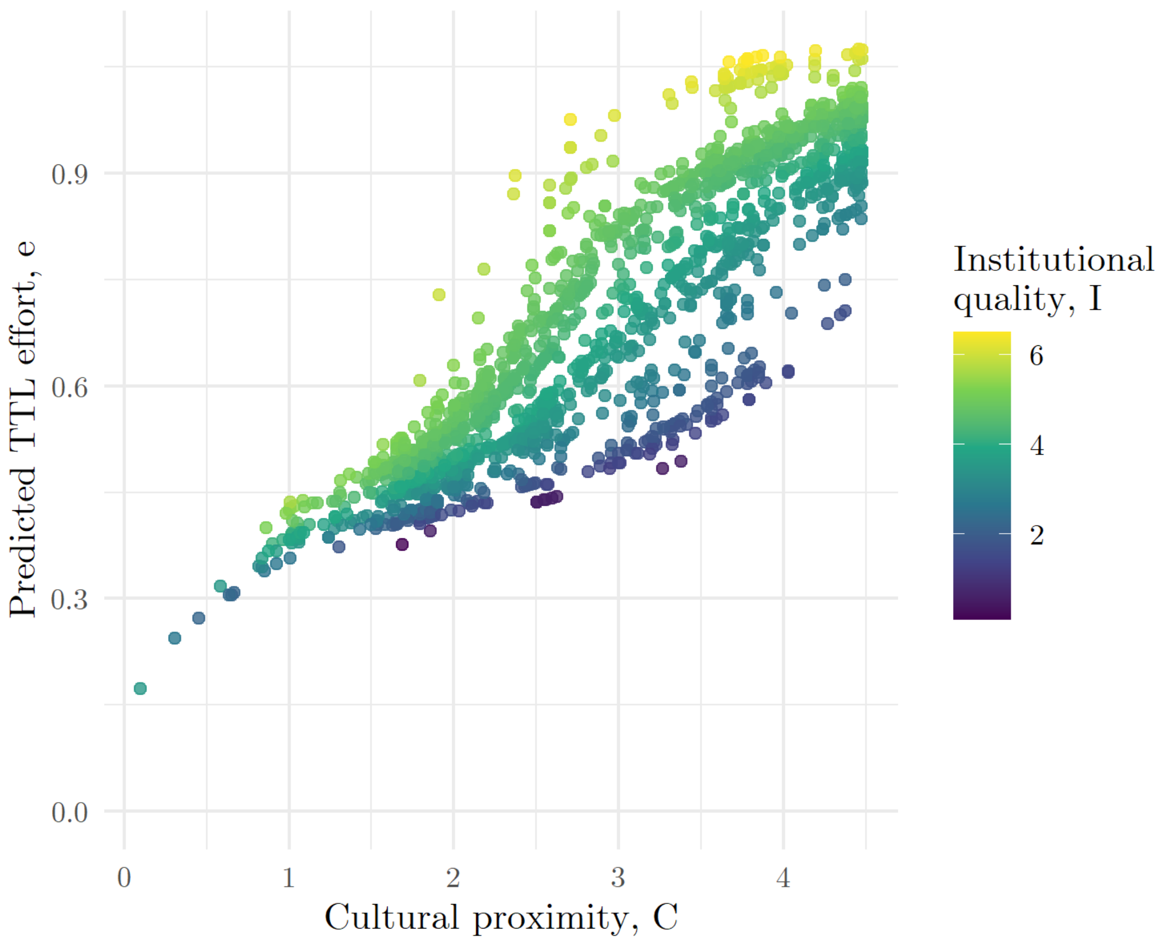
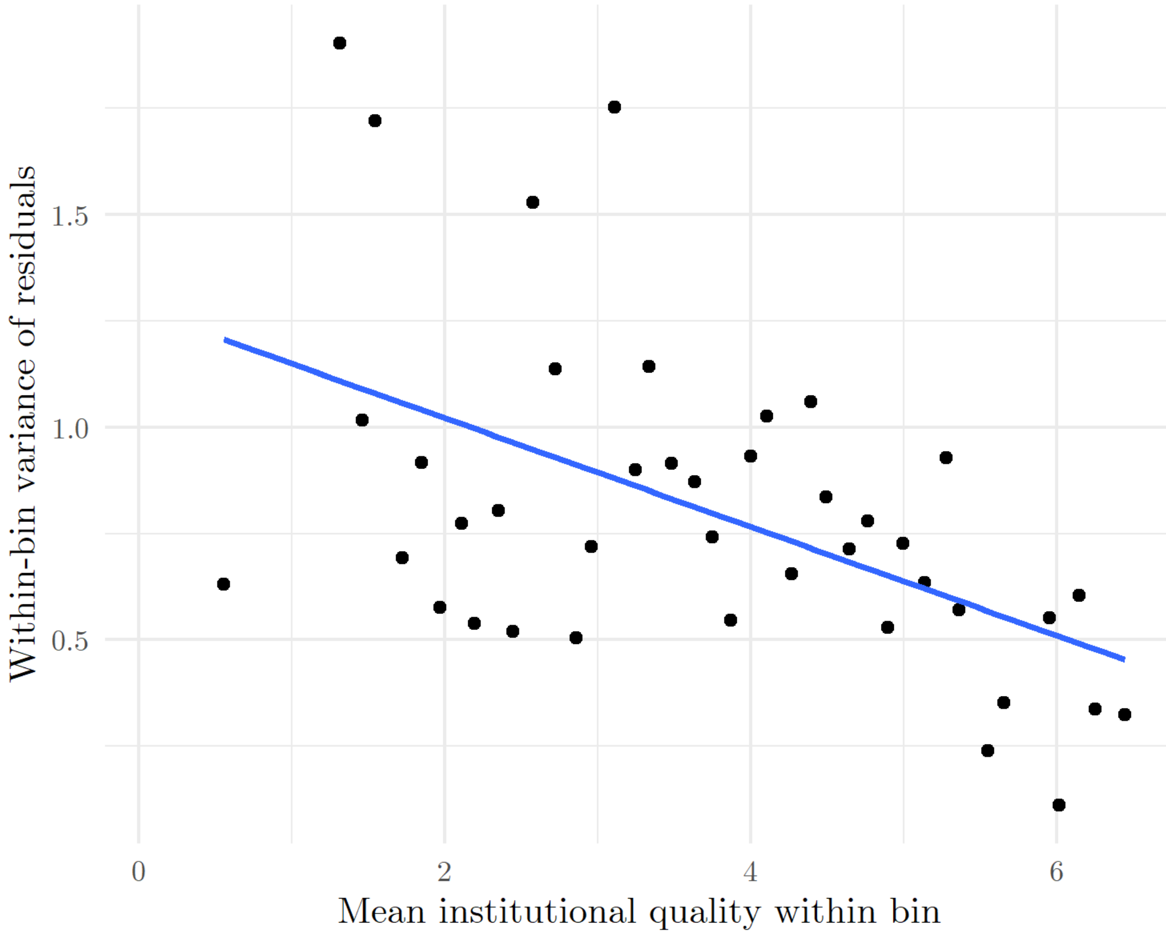


Figure 4: Fitting $h(I)$



C Instrument validation

Table C1 presents estimation results for tests of the IV's exclusion restriction as discussed in Section 5. The instrument, the average cultural proximity to each recipient country of TTLs who led any World Bank project that was evaluated before 2005, is intended to capture the availability of culturally close TTLs to each recipient country.

Table C1: Tests of the IV's exclusion restriction

	Dependent variable is the instrument		
	(1)	(2)	(3)
Number of projects	−0.048		
(in hundreds)	(0.406)		
Total project costs		−0.047	
(in 10s of billions)		(0.115)	
CPIA			0.118
			(0.235)
log(population)	0.284***	0.297***	0.269***
	(0.097)	(0.084)	(0.066)
log(GDP)	0.393**	0.406**	0.359**
	(0.160)	(0.160)	(0.164)
Observations	75	75	75
R^2	0.199	0.200	0.201
Note:		*p<0.1; **p<0.05; ***p<0.01	

Instrument is average distance to each recipient country of TTLs who led pre-sample projects.

D Surnames for female TTLs

Section 6 evaluates the possibility that cultural proximity is systematically mismeasured for female TTLs by comparing my primary measure of cultural proximity with an alternative measure of cultural proximity that should not vary in accuracy between male and female TTLs. Another way to evaluate the possibility that cultural proximity is systematically mismeasured for female TTLs is to allow the key coefficients in my model to vary by gender. Table D1 reports results from two IV specifications with key coefficients interacted with an indicator variable equal to 1 if the TTL is female. Model 1 in Table D1 reports coefficients estimated by a model including the full set of interaction terms between *Cultural proximity*, *Institutional quality*, and *Female*. The coefficient on *Cultural proximity*, which is now identified only by variation among male TTLs, is qualitatively similar to the estimate from my preferred specification (model 2 in Table 4), but the key interaction term *Cultural proximity* $\times$ *Institutional quality* is smaller and not statistically distinguishable from 0.

However, none of *Cultural proximity* $\times$ *Female*, *Institutional quality* $\times$ *Female*, or *Cultural proximity* $\times$ *Institutional quality* $\times$ *Female* has a statistically significant effect on project success. Furthermore, these three variables are not jointly statistically significant ($F = 1.134$, $p = 0.334$). In short, we have no evidence that the gender of a project's TTL is an important determinant for the success of the project, either on its own or in conjunction with *Cultural proximity* or *Institutional quality*.

Table D1: IV results with gender controls, selected coefficients

	IEG Outcome: 1-6 scale		
	(1)	(2)	(3)
Cultural proximity	0.215^{*}	0.216^{*}	0.233^{*}
	(0.127)	(0.126)	(0.125)
Institutional quality	0.204^{**}	0.265^{***}	0.267^{***}
	(0.097)	(0.079)	(0.075)
Female	0.085	0.119	-0.021
	(0.140)	(0.142)	(0.053)
Cultural proximity ×	0.074	0.114^{**}	0.111^{**}
Institutional quality	(0.061)	(0.047)	(0.046)
Cultural proximity ×	0.091	0.101	
Female	(0.097)	(0.099)	
Institutional quality ×	0.179	0.017	
Female	(0.127)	(0.054)	
Cultural proximity ×	0.109		
Institutional quality ×	(0.079)		
Female			

Notes: For a summary of additional control variables included in all models, see notes to Table 2. Twenty-one observations for which TTL gender could not be determined have been excluded.

E Alternative clustering and robustness checks

Tables E1, E2 and E3 are analogous to Table 4, but allow standard errors to cluster at alternative levels of aggregation. Table E1 presents estimates with standard errors clustered at the country level. Table E2 presents estimates with standard errors clustered at the level of country-project approval year. Finally, Table E3 presents estimates with unclustered standard errors. In all cases, the coefficient estimates for *Cultural proximity* $\times$ *Institutional quality* are statistically significant from 0 at the 5% level. I do not cluster standard errors by only approval year or only project evaluation year. These levels of aggregation yield only 20 and 14 clusters, respectively. As a rule of thumb, fewer than 40 clusters is not sufficient to estimate clustered standard errors.

Table E4 presents OLS results with controls for country fixed effects. The inclusion of country fixed effects precludes use of the instrument, which varies only between countries. Institutional quality has an insignificant effect on project outcomes after including country fixed effects because there is little variation in institutional quality within a country over time. The direct effect of cultural proximity remains significant at the 5% level in specification (1), but the interaction of cultural proximity and institutional quality is insignificant.

Table E5 presents IV results for a series of linear probability models, where the cut off for "success" is assigned to every possible value. That is, in column (1), the outcome variable is 1 if the project received an IEG score greater than 1. In column (2), the outcome variable is 1 if the project received an IEG score greater than 2, and so on. Together, these specifications demonstrate that cultural proximity is especially important for producing projects that receive the highest IEG outcome score of 6. Additionally, cultural proximity and its interaction with institutional quality are especially important for achieving at least an IEG outcome of 4. On the other hand, cultural proximity plays little role in saving projects from the worst scores of 1 and 2.

Figure 5 plots the coefficient estimates from six linear probability models, where each dependent variable is coded as 1 if a project received score i, $i \in 1, 6$. A 1-σ increase in cultural proximity (above the mean), for instance, is associated with a significantly lower probability of receiving an IEG score of 3, but is associated with a significantly higher probability of receiving a score of 6. None of the variables—institutional quality, cultural proximity, or the interaction of the two—significantly changes a project's probability of receiving a score of 1 or 2. In general terms, better institutions and more cultural proximity are associated with fewer marginal failures (scores of 3) but not with fewer abject failures (scores of 1 or 2).

Table E6 presents IV results using only a subset of the CPIA composite score as a measure of institutional quality. The World Bank constructs its CPIA score as a simple average of four cluster subscores for economic management, structural policies, social inclusion/equity, and public sector management. Using the simple average of these as a proxy for overall institutional quality may overstate or understate the relative importance of each component.

Table E6 reproduces the results from the main IV specification (model 2 in Table 4), using each CPIA cluster subscore in place of the single CPIA aggregate. The results indicate that social and economic institutions have the strongest interaction with cultural proximity.

Figure 5: Effect of selected variables on receiving specific IEG scores

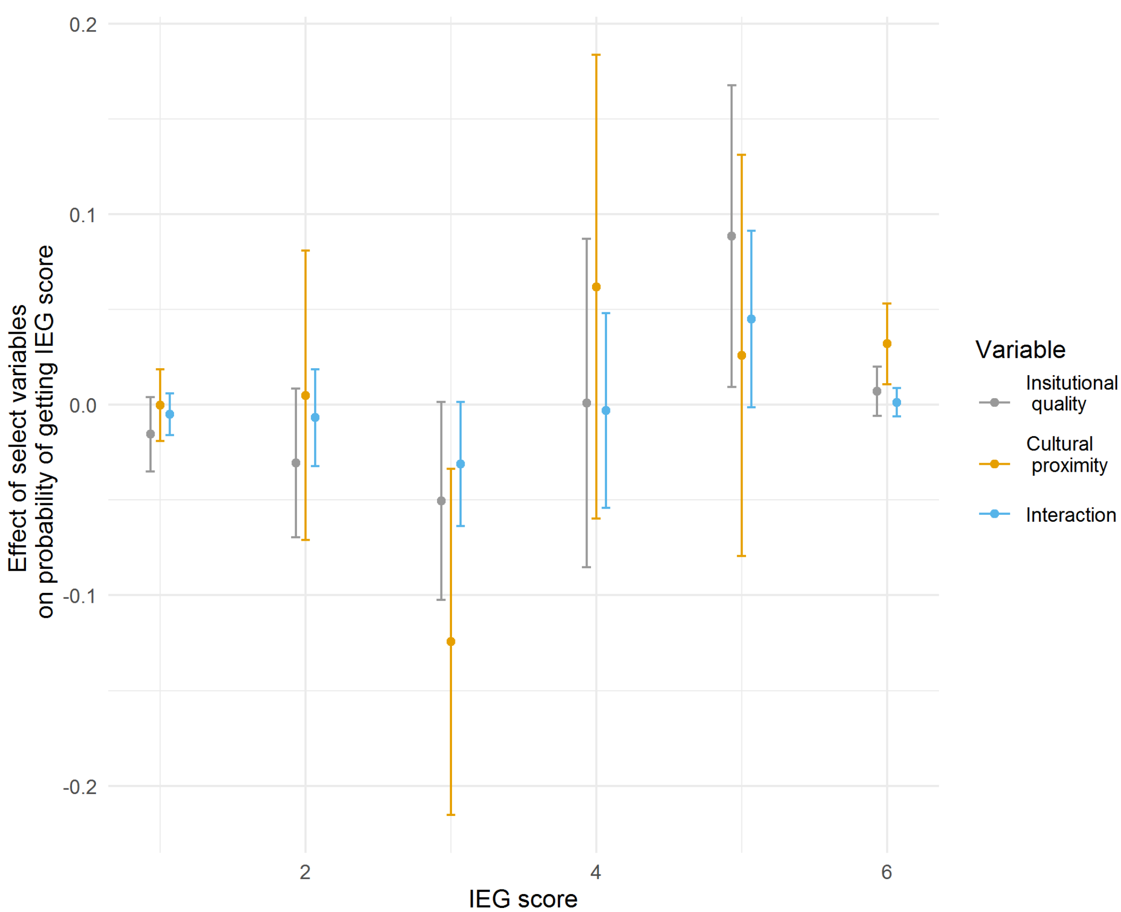


Table E1: IV results, standard errors clustered by country

	Dependent variable: Project success		
	(1)	(2)	(3)
Cultural proximity	0.185	0.205^{*}	0.185
	(0.128)	(0.123)	(0.123)
Institutional quality	0.095^{***}	0.261^{***}	0.313^{***}
	(0.036)	(0.080)	(0.101)
Cultural proximity ×		0.107^{**}	0.130^{**}
Institutional quality		(0.045)	(0.055)
Other controls	Yes	Yes	Yes
Reviewer FEs	No	No	Yes
N	1,946	1,946	1,946

Notes: $^{***} < .01$, $^{**} < .05$, $^{*} < .1$

For a summary of additional control variables included in all models, see notes to Table 2.

Table E2: IV results, standard errors clustered by country-approval year

	Dependent variable: Project success		
	(1)	(2)	(3)
Cultural proximity	0.185	0.205^{*}	0.185
	(0.118)	(0.118)	(0.121)
Institutional quality	0.095^{***}	0.261^{***}	0.313^{***}
	(0.030)	(0.075)	(0.079)
Cultural proximity ×		0.107^{**}	0.130^{***}
Institutional quality		(0.044)	(0.047)
Other controls	Yes	Yes	Yes
Reviewer FEs	No	No	Yes
N	1,946	1,946	1,946
Notes:			$^{***} < .01$, $^{**} < .05$, $^{*} < .1$

For a summary of additional control variables included in all models, see notes to Table 2.

Table E3: IV results, standard errors unclustered

	Dependent variable: Project success		
	(1)	(2)	(3)
Cultural proximity	0.185*	0.205*	0.185*
	(0.111)	(0.113)	(0.112)
Institutional quality	0.095***	0.261***	0.313***
	(0.026)	(0.071)	(0.073)
Cultural proximity ×		0.107**	0.130***
Institutional quality		(0.043)	(0.044)
Other controls	Yes	Yes	Yes
Reviewer FEs	No	No	Yes
N	1,946	1,946	1,946

Notes: $^{***} < .01$, $^{**} < .05$, $^{*} < .1$

For a summary of additional control variables included in all models, see notes to Table 2.

Table E4: OLS results with country fixed effects

	Dependent variable: Project success		
	(1)	(2)	(3)
Cultural proximity	0.067**	0.063*	0.063
	(0.034)	(0.034)	(0.038)
Institutional quality	−0.067	−0.009	0.068
	(0.125)	(0.132)	(0.150)
Cultural proximity ×		0.044	0.028
Institutional quality		(0.030)	(0.034)
Country FEs	Yes	Yes	Yes
Other controls	Yes	Yes	Yes
Reviewer FEs	No	No	Yes
N	1,946	1,946	1,946

Notes: $^{***} < .01$, $^{**} < .05$, $^{*} < .1$

All models include controls listed in the notes to Table 2, with the exception of the geographic region of recipient countries, which are subsumed by country fixed effects.

Table E5: IV linear probability models

	Project success = 1 if...				
	IEG score > 1	IEG score > 2	IEG score > 3	IEG score > 4	IEG score > 5
	(1)	(2)	(3)	(4)	(5)
Cultural proximity	0.000	-0.005	0.120**	0.058	0.032***
	(0.010)	(0.039)	(0.057)	(0.055)	(0.011)
Institutional quality	0.015	0.046**	0.096***	0.096**	0.007
	(0.010)	(0.019)	(0.033)	(0.042)	(0.007)
Cultural proximity ×	0.005	0.012	0.043**	0.046*	0.001
Institutional quality	(0.006)	(0.013)	(0.021)	(0.024)	(0.004)

Notes: $^{***} < .01$, $^{**} < .05$, $^{*} < .1$

For a summary of additional control variables included in all models, see notes to Table 2.

Table E6: IV Results for CPIA clusters

	IEG Outcome: 1-6 scale			
CPIA cluster:	Social inclusion (1)	Economic management (2)	Public sector (3)	Structural policies (4)
Cultural proximity	0.186	0.193	0.194	0.140
	(0.135)	(0.124)	(0.121)	(0.123)
CPIA cluster	0.256^{***}	0.236^{***}	0.234^{***}	0.157^{*}
	(0.087)	(0.070)	(0.064)	(0.085)
Cultural proximity × CPIA cluster	0.114^{**}	0.108^{**}	0.082^{*}	0.067
	(0.055)	(0.046)	(0.047)	(0.059)
N	1841	1946	1946	1946

Notes: $^{***} < .01$, $^{**} < .05$, $^{*} < .1$

CPIA cluster scores are standardized $\mu = 0$, $\sigma = 1$. Social inclusion cluster scores are missing for projects that took place in Indonesia (105 observations). For a summary of additional control variables included in all models, see notes to Table 2.

F Dyadic regression results

Dyadic regressions are used to model network connectivity. Often, dyadic regressions model the formation of links between nodes, where any two nodes may share a link. Every pair of nodes is an observation, and the outcome variable is equal to 1 for pairs that share a link, 0 otherwise. Explanatory variables can be features of one or both nodes, or of the relation between them. Here I use dyadic regressions to model the probability that TTLs are assigned to projects. In this case, an observation is a TTL-project pair, and all possible TTL-project pairs are observations. Note that this differs from the typical case in that TTL-TTL and project-project pairs are not possible. The outcome variable is equal to 1 if that TTL was the primary TTL for the project. Regressing the outcome variable on the cultural distance between the TTL and the project's recipient country measures the role of cultural proximity in the assignment process of TTLs to projects. The results in Table F1 indicate that TTLs who are culturally close to a recipient country are more likely to be assigned to a project in that country. Additionally, this tendency is somewhat stronger for recipient countries with strong institutions, though this relationship is significant only at the 10% level. Taken together, these results suggest that the World Bank's TTL assignment policies already conform with the recommendation that culturally close TTLs should be preferentially assigned to projects, but that World Bank projects may benefit further if this assignment policy is emphasized in countries with strong institutions.

Table F1: Dyadic regressions

	Dependent variable:		
	= 1 if TTL_i is assigned to project_j		
	(1)	(2)	(3)
Cultural proximity	0.550***	0.555***	0.552***
	(0.022)	(0.022)	(0.022)
Institutional quality		−0.052***	0.021
		(0.009)	(0.041)
Cultural proximity × Institutional quality			0.033*
			(0.018)
Observations	2,247,023	2,247,023	2,247,023

Note: *p<0.1; **p<0.05; ***p<0.01

All models include TTL fixed effects. Standard errors are clustered at the project level. Coefficients and standard errors are multiplied by 1000 for readability.

References

Alesina, Alberto et al. “Fractionalization”. In: *Journal of Economic growth* 8.2 (2003), pp. 155–194.

Arndt, Channing, Sam Jones, and Finn Tarp. “Aid, growth, and development: have we come full circle?” In: *Journal of Globalization and Development* 1.2 (2010).

Burnside, Craig and David Dollar. “Aid, policies, and growth”. In: *American economic review* 90.4 (2000), pp. 847–868.

Cassen, Robert et al. “Does aid work?: report to an intergovernmental task force”. In: *OUP Catalogue* (1994).

Chauvet, Lisa, Paul Collier, and Andreas Fuster. “Supervision and project performance: A principal-agent approach”. In: (2017).

Denizer, Cevdet, Daniel Kaufmann, and Aart Kraay. “Good countries or good projects? Macro and micro correlates of World Bank project performance”. In: *Journal of Development Economics* 105 (2013), pp. 288–302.

Diallo, Amadou and Denis Thuillier. “The success of international development projects, trust and communication: an African perspective”. In: *International journal of project management* 23.3 (2005), pp. 237–252.

Dollar, David and Victoria Levin. *Sowing and reaping: institutional quality and project outcomes in developing countries*. The World Bank, 2005.

Dollar, David and Jakob Svensson. “What explains the success or failure of structural adjustment programmes?” In: *The Economic Journal* 110.466 (2000), pp. 894–917.

Doucouliagos, Hristos and Martin Paldam. “The aid effectiveness literature: The sad results of 40 years of research”. In: *Journal of Economic Surveys* 23.3 (2009), pp. 433–461.

— “The ineffectiveness of development aid on growth: An update”. In: *European journal of political economy* 27.2 (2011), pp. 399–404.

Duponchel, Marguerite, Lisa Chauvet, and Paul Collier. *What explains aid project success in post-conflict situations?* The World Bank, 2010.

Easterly, William. “Can foreign aid buy growth?” In: *Journal of economic Perspectives* 17.3 (2003), pp. 23–48.

Geli, Patricia, Aart Kraay, and Hoveida Nobakht. *Predicting world bank project outcome ratings*. The World Bank, 2014.

Guillaumont, Patrick and Rachid Laajaj. *When instability increases the effectiveness of aid projects*. The World Bank, 2006.

Ika, Lavagnon A, Amadou Diallo, and Denis Thuillier. "Critical success factors for World Bank projects: An empirical investigation". In: *International journal of project management* 30.1 (2012), pp. 105–116.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. "The worldwide governance indicators: methodology and analytical issues". In: *Hague Journal on the Rule of Law* 3.2 (2011), pp. 220–246.

Kilby, Christopher. "Supervision and performance: the case of World Bank projects". In: *Journal of Development Economics* 62.1 (2000), pp. 233–259.

Luoto, Jill et al. "What point-of-use water treatment products do consumers use? Evidence from a randomized controlled trial among the urban poor in Bangladesh". In: *PloS one* 6.10 (2011), e26132.

Mold, Andrew. "Will it all end in tears? Infrastructure spending and African development in historical perspective". In: *Journal of International Development* 24.2 (2012), pp. 237–254.

Nguyen, Hai H, Alexandros Alexiou, and Alexander Singleton. "Personal Name Classification Using Collective Data". In: (2017).

Pemberton, Trevor J, Michael DeGiorgio, and Noah A Rosenberg. "Population structure in a comprehensive genomic data set on human microsatellite variation". In: *G3: Genes, Genomes, Genetics* 3.5 (2013), pp. 891–907.

Pursiainen, Vesa. "Cultural Biases in Equity Analysis". In: *Available at SSRN 3153900* (2019).

Qian, Nancy. "Making progress on foreign aid". In: *Annu. Rev. Econ.* 7.1 (2015), pp. 277–308.

Rajan, Raghuram G and Arvind Subramanian. "Aid and growth: What does the cross-country evidence really show?" In: *The Review of economics and Statistics* 90.4 (2008), pp. 643–665.

Spolaore, Enrico and Romain Wacziarg. "Ancestry and development: New evidence". In: *Journal of Applied Econometrics* 33.5 (2018), pp. 748–762.

— "Ancestry, language and culture". In: *The Palgrave handbook of economics and language*. Springer, 2016, pp. 174–211.

Spolaore, Enrico and Romain Wacziarg. "The diffusion of development". In: *The Quarterly journal of economics* 124.2 (2009), pp. 469–529.

Temple, Jonathan RW. "Aid and conditionality". In: *Handbook of development economics*. Vol. 5. Elsevier, 2010, pp. 4415–4523.