Cultural Constraints on Managerial Productivity

Evidence from World Bank Lending

Yves Froude †

Aaditya Mattoo†

Bob Rijkers§ †

Michele Valsecchi‡ †

February 2026

Abstract: Does cultural misalignment between project managers and the environments in which they operate undermine their effectiveness? Using administrative and project-level data from the World Bank and exploiting rotations of the same managers across countries, we show that project outcome ratings decline when managers' visible attributes conflict with prevailing norms: projects led by women receive lower ratings in more gender-unequal environments, and projects led by younger managers receive lower ratings in settings where seniority is highly valued. By contrast, less conspicuous forms of cultural distance, such as nationality-based measures, do not predict project outcomes. We find no evidence that these effects are driven by sorting on managerial ability. The results are robust to controlling for country and manager fixed effects and to using an instrumental-variables strategy exploiting variation in the supply of eligible managers to address endogenous assignment. Back-of-the-envelope calculations suggest that the costs of cultural misalignment are approximately 1.5 percent of total aid flows.

Key words: gender, gender inequality, gerontocracy, cultural distances, aid effectiveness, development

JEL codes: O10, O16, O22, J16, D73

§ Utrecht University, † Development Research Group, World Bank, ‡ New Economic School (NES) and University of Gothenburg.
We thank Alina Timoshkina, Cung Truong Huoang and Natalia Pecorari for excellent research assistance, Daria Tagloni, Jed Friedman, and Clement Joubert for excellent comments and help accessing the data. We also benefited from discussions with the leadership team from the Human Resources Vice Presidency of the World Bank, and from comments from and discussions with Radi Anguelova, S. Anukriti, Reena Badiani, Kathleen Beegle, Emmet Cunningham, Louise-Paul Delvaux, John Giles, Seema Jayachandran, Eeshani Kandpal, Tatjana Kleineberg, and Joanne Marie McDonough, David McKenzie, Nathan Nunn, Hosny Zoabi, as well as seminar participants of the ICEF workshop on Applied Economics, the University of Gothenburg, the Jobs Workshop at Naas Fabriken, and the World Bank Poverty and Applied Microeconomics Seminar series. This paper has been partly supported by the Umbrella Facility for Trade trust fund (financed by the governments of the Netherlands, Norway, Sweden, Switzerland, and the United Kingdom) and the Strategic Research Partnership on Economic Develop+ment. We also acknowledge the generous financial support from the World Bank research support budget. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank of Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the countries they represent. All errors are ours. The authors can be contacted at yfroude@worldbankgroup.org, mvalsecchi@worldbank.org, amattoo@worldbank.org and brijkers@worldbank.org.

1 Introduction

Why do the same professionals performing similar tasks achieve systematically different outcomes across countries? Alignment with local cultural norms may provide part of the answer. Individuals whose attributes are less favored by local norms may achieve lower performance, even when skills and incentives are identical. Such mismatches may be particularly consequential for development projects, since the professionals tasked with executing them often operate in cultural contexts that differ sharply from their own.

Identifying the impact of culture on productivity is challenging. Professionals sort into roles and environments and differ systematically in both observable and unobservable characteristics, complicating causal inference. It is rarely possible to observe the same individual performing comparable tasks across multiple cultural contexts. Moreover, performance is often difficult to measure, and observed differences may reflect variation in incentives, objectives, or institutional conditions rather than productivity itself.

We make progress on these challenges by leveraging a unique setting. We study World Bank project managers—Task Team Leaders (TTLs)—who are directly accountable for project outcomes and rotate across recipient countries while operating under a standardized organizational framework. TTLs face common incentives and perform comparable tasks, yet work in environments that vary widely in local cultural norms. Using three decades of matched project and personnel data, we follow the same TTLs across countries, enabling within-manager comparisons that hold ability and incentives constant.

Using this setting, we test whether cultural misalignment constrains the effectiveness of foreign aid, with a particular focus on gender and age. Specifically, we examine whether (i) projects led by women perform differently than those led by men in countries with higher gender inequality, and (ii) projects led by younger TTLs perform differently than those led by older TTLs in societies that place greater value on seniority. We also examine whether project performance varies with cultural distance between a TTL's nationality and the recipient country, along multiple dimensions including historical ties (past conflict or colonial links), linguistic distance, and differences in cultural values, such as collectivism or attitudes toward work–life balance, elicited in surveys of TTLs conducted during cultural awareness trainings. Exploiting the rotation of TTLs across recipient countries, we compare the same manager operating in different cultural environments, isolating the interaction between manager attributes and local norms while netting out time-invariant individual heterogeneity and task-level differences. We further address selection into project leadership by leveraging plausibly exogenous variation in the gender and age composition of the pool of potential TTLs.

Our main findings can be summarized as follows. First, we document systematic gender sorting into project leadership roles that correlates with local cultural norms. Women are significantly less likely to lead projects in countries with more gender-unequal norms. By contrast, we find little evidence that younger TTLs sort based on local norms favoring seniority.

Importantly, we find no evidence of systematic sorting by ability. Women (and men) leading projects in gender-unequal countries are not, on average, less experienced, educated, or capable than TTLs operating in more gender-equal environments. Similarly, younger TTLs assigned to countries where norms favor the elderly are as capable as those working

countries where norms regard youth more favorably.

Second, and most important, using a difference-in-difference design, we show that cultural misalignment compromises development effectiveness. Development projects led by female TTLs suffer a performance penalty in gender-unequal countries relative to those led by men. Similarly, projects led by young TTLs suffer a performance penalty relative to those led by older TTLs in countries with more favorable attitudes towards the elderly. By contrast, we cannot reject the null hypothesis that less salient forms of cultural distance such as historical conflict between the TTL's and recipient countries or misalignment in cultural values balance, predict project outcomes. These result are robust to controlling for TTL and country fixed effects, further assuaging concerns that the result reflects sorting based on ability of the TTL and/or unobserved country characteristics.

These results are robust to a variety of tests, including addressing potential residual selection bias with an instrumental variable strategy, that exploits variation in the gender and age composition of the pool of potential project leaders, notably staff at the same grade level in the unit responsible for the project. Results also obtain using alternative proxies for attitudes towards the elderly and alternative measures of gender norms. Finally, we provide suggestive evidence that the observed performance penalties are not driven by reductions in observable inputs. We conduct analogous analyses for two key inputs into a TTL's production function—the number of trips to the project country and the use of temporary workers (short-term consultants). We find no systematic differences along either margin, which is consistent with the interpretation that the effects on project outcomes do not primarily operate through reduced effort or resource use.

How costly are these cultural constraints? A back-of-the-envelope calculation reported in the Appendix suggests that, under current TTL assignments, project success rates would be 3.2% higher if all recipient countries had levels of gender inequality equal to those at the 10th percentile of the global distribution, 0.5% higher if all countries had attitudes toward seniority at the 10th percentile. Put differently, absent cultural misalignment, the World Bank could have achieved the same development outcomes with approximately 1.38% less spending. These cultural constraints also contribute to gender and age pay gaps, as young and female project managers operating in culturally hostile environments receive lower performance ratings and, consequently, smaller pay increases.

This paper builds on and contributes to several strands of work on gender and development. Prior research shows that gender inequality hampers growth, inter alia by erecting barriers to capital accumulation (Klasen, 2018) and entrepreneurship (Chiplunkar and Goldberg, 2024), and inducing misallocation of talent (Hsieh et al., 2019). Yet, direct cross-country evidence on the role of norms remains scarce. A notable exception is Ashraf et al. (2024), who also exploit micro-level data from a multinational organization, but address a distinct question. Their focus is on how women self-select into the firm as a function of local labor market conditions, explaining variation in productivity across individuals. In contrast, we study how the same women vary their productivity within individuals across projects, depending on the gender norms prevailing in the project location.

Second, and related, our project contributes to the cultural economics literature. Several papers have shown that failing to take into consideration local norms can undermine policy effectiveness: medical campaigns may face resistance (Lowes and Montero, 2021); agricultural interventions can backfire (Lansing, 2009) or even fuel conflict (McGuirk and

Nunn, 2024); land titling programs may prove ineffective (Le Rossignol et al., 2024); and social welfare programs can generate spillovers to non-eligible households (Moscona and Seck, 2024) (See Bau et al. (2025) for a recent review of this literature). Leveraging global data and within-manager variation, we show that the same individuals perform differently across cultural environments, allowing us to separate external norms that vary by location from internal norms carried by managers. In contrast to work emphasizing cultural distance alone (e.g., Ek (2024) and Nguyen (2018)), we highlight interactions between demographic traits and local norms.

Third, building on Saenger (2023) and Mulligan and Sala-i Martin (1999) we are among the first few papers to examine the consequences of gerontocracy.

Fourth, we contribute to the literature on aid effectiveness in general (for reviews, see Qian (2015); Arndt et al. (2016)), and the success of individual projects in particular. Previous studies have shown that project manager characteristics, project preparation, and recipient country factors partly predict project outcomes (Kilby (2015), Denizer et al. (2013), Ashton et al. (2023), Geli et al. (2014)), with project manager characteristics being especially important (Moscona (2025)). Building on this, we provide the first systematic evidence quantifying how the gender and age of the project leader interacts with local norms to influence project outcomes, showing that local culture shapes project performance. Our findings indicate that, along with institutional quality (Burnside and Dollar (2000, 2004)), gender and age based cultural fit is a measurable and economically meaningful determinant of project success.

Structure of paper The remainder of this paper is organized as follows. The next section describes the institutional context and the data. Section 3 describes our identification strategy. Results are presented in section 4. A final section concludes.

2 Institutional Context

The World Bank is the world's largest multilateral development lender, financing operations in more than 145 countries. Each project is managed by one or more Task Team Leaders (TTLs), who are responsible for project design, supervision, and outcomes, and who serve as primary interlocutors with recipient governments. To isolate individual contributions, our main analysis excludes projects jointly managed by multiple TTLs.

All TTLs operate within a common institutional framework: projects must follow standardized preparation, supervision, and evaluation procedures, and outcomes are assessed using harmonized metrics, including ratings from the Bank's Independent Evaluation Group (IEG). As its name suggest, the IEG is independent from World Bank management and provides an autonomous validation of project ratings. This consistency allows project performance to be compared across countries while holding institutional rules constant.

TTL assignments are made by managers based on operational needs, technical expertise, and staff availability. Junior staff (grades GE–GF) typically manage smaller or less risky projects, while senior specialists and leads (GG–GH) take on larger, technically demanding, or politically sensitive operations. Individual staff preferences and career progression are also taken into consideration. Serving as TTL is widely regarded as a desirable role, providing

visibility within the organization and potential career rewards.

The World Bank offers a suitable setting to study the impact of gender inequality on development effectiveness. First, since many TTLs work in multiple countries, we can observe the same individual operating in different cultural contexts while holding institutional processes and incentives constant. Second, we can observe many TTL characteristics, which allows us to test directly for selection. Third, there is substantial variation in the share of staff with specific across sectors and grades, as we will discuss below. This creates systematic variation in the pool of potential young and female TTLs across units and grades, which we exploit in our identification strategy. Last but not least, the World Bank's explicit commitments and policies to promote equity and inclusion, together with its active efforts to identify and remedy discriminatory practices, make it unlikely that observed differences in TTL performance simply reflect institutional bias.

3 Data

The project combines various datasets:

World Bank Project Data We construct a dataset of World Bank projects implemented between 1987 and 2023, combining multiple sources of information in three steps. First, we collect standard project-level data from the World Bank project portal, including approval date, expected and actual duration, sector, recipient country, and total amount committed (in USD).

Second, and most novel, we manually compile information on Task Team Leaders (TTLs). For each project, we identify the TTL at approval ("Designer TTL") and at completion ("Supervisor TTL"), drawing primarily from Implementation Completion Reports (ICRs), the World Bank's mandatory end-of-project self-evaluation documents. When ICRs are unavailable or incomplete, we supplement the data with intermediate project documents and Project Appraisal Documents (PADs), which are prepared by Bank staff just before consideration by the World Bank's Executive Board.

Third, we merge these data with outcome ratings from the World Bank's Independent Evaluation Group (IEG), which evaluates completed projects on a standardized six-point scale ranging from Highly Unsatisfactory (1) to Highly Satisfactory (6). Projects rated 1–3 are classified as unsuccessful, while those rated 4 or higher are considered successful. These ratings provide a comparable measure of project performance across countries. Denizer et al. (2013).[1]

In our analysis, we restrict the sample to IDA- and IBRD-financed investment projects with published IEG outcome ratings, known TTLs, and executed in a single country.[2] Our

[1]They have been shown to correlate with objective outcomes. for example Andersen et al. (2022) show that aid disbursements to projects that fail in highly aid-dependent countries are associated with increases in offshore wealth held in tax havens, whereas disbursements to successful projects are not.

[2]We exclude development policy loans, projects financed by trust funds, multisectoral projects, and projects spanning multiple countries. Our starting dataset contains 7,955 projects with known IEG outcome ratings, of which 220 cannot be assigned to a unique country and 15 lack information on the main sector. We further exclude 757 development policy loans, 102 projects for which we cannot identify the relevant organizational unit of the TTL in the HR data, and 1,129 projects with multiple supervisor TTLs.

main sample therefore consists of 4,367 projects with a single supervisor TTL.

Human Resources Data We use administrative human resources (HR) data of World Bank staff records for the period 1987-2024. These data include information on staff gender, nationality, age, education, family composition, date of entry, position, grade, salary, contract type, duty station, and organizational unit. In addition, we observe annual performance ratings for most staff, scored on a scale from 1 (totally unsatisfactory) to 5 (superb), with the vast majority receiving a 3. Higher ratings are associated with faster salary progression.

Travel data We also obtain data on World Bank staff travel patterns for the period 2000-2024. The data include information on the date, duration, destinations and purpose of the trip.

Gender inequality We use four commonly used cross-country indicators of gender inequality as proxies for gender norms. First, the Global Gender Gap Index (GGGI), published by the World Economic Forum (WEF), measures gender gaps across economic participation, educational attainment, health, and political empowerment (Forum, 2021). Second, the Social Institutions and Gender Index (SIGI), compiled by the OECD Development Centre, measures discrimination embedded in social institutions across four domains: discrimination in the family, restricted physical integrity, restricted access to productive and financial resources, and restricted civil liberties (Centre, 2023). Third, the Gender Inequality Index (GII), published by the United Nations Development Programme (UNDP), reflects gender-based disadvantages in reproductive health, empowerment, and labor market participation (Programme, 2020). Finally the Women, Business and the Law (WBL) Index, produced by the World Bank, captures legal barriers to women's economic opportunities across ten domains covering the lifecycle of an economically active woman: safety, mobility, workplace, pay, marriage, parenthood, childcare, entrepreneurship, assets, and pensions (Bank, 2023). The latter is a measure of *de jure* discrimination that may not capture reality on the ground. We take country-level averages over the sample period for each index, producing time-invariant measures that better capture longstanding gender norms rather than short-term fluctuations. All four indices are standardized to be have a mean of zero and a standard deviation of one, and rescaled such that higher values indicate greater gender inequality. They are highly but not perfectly correlated, with correlations hovering between 0.43 and 0.71.

Attitudes towards the elderly To measure attitudes towards the elderly we use data on implicit bias against the elderly from Project Implicit, which conducts free online Implicit Association Tests (IAT). These measure strengths of associations between concepts (e.g. "the young" vs "the elderly") and evaluations (e.g. "good" vs "bad"), establishing measures of bias by comparing reactions times across categorization tasks. Positive bias towards the elderly is inferred when survey participants associate older people more quickly with positive attributes and younger people with negative attributes, compared to the reverse pairing. The IAT is designed to capture automatic cognitive responses shaped by cultural

and social norms, and thereby to limit the impact of social desirability bias. We also use similar data to measure implicit attitudes towards women. One limitation of these measures is that the sample of respondents is self-selected, and hence not necessarily representative of the underlying population. However, Figure 5 shows that our IAT-based measure correlate reasonably well with survey-based measures of attitudes towards the elderly from the World Values Survey, which are available only for a subset of World Bank client countries.

In robustness checks, we also use an alternative measure of attitudes towards the elderly developed by Saenger (2023), who builds on the folklore motif approach developed by Michalopoulos and Xue (2021), which uses ethnographic records of traditional narratives to capture persistent cultural norms. These motifs capture traditional beliefs about the social role and status of older individuals and are highly persistent over time. Saenger (2023) shows that countries characterized by more positive elderly-related motifs tend to have older political leaders and more favorable contemporary attitudes toward the elderly, consistent with this measure capturing enduring gerontocratic norms.

Cultural traits To measure cultural traits at the individual level, we use data collected by Culture Wizard (CW), a firm specializing in the measurement of cultural attitudes and diversity training. The CW survey elicits employees' views along eight dimensions: Importance of Work, Change–Control–Risk, Collectivism, Hierarchy, Time Management, Transactional Orientation, Direct Communication, and Informal Communication. More than 10,000 World Bank employees completed the CW survey.

In addition to these individual-level measures, we use several proxies for cultural traits based on staff members' country of origin. These include country-level measures derived from Culture Wizard, as well as standard measures of cultural distance commonly used in the literature. Specifically, we include (i) an indicator for whether the project manager's country of origin shares an official language with the recipient country, (ii) ethnolinguistic distance, and (iii) a measure of cultural distance developed by Pellegrino, Spolaore and Wacziarg (2025). Finally, for a subset of staff, we also observe individual cultural views.

Other data We supplement these data with information on Gross Domestic Product (GDP) per capita from the World Development Indicators (WDI), wars and conflict from the Upssala Conflict Data Program, and financial and debt crises from Leaven and Valencia (2020).

4 Empirical Strategy

This study examines whether the effectiveness of World Bank projects depends on the alignment between leader characteristics and prevailing social norms in recipient countries. Specifically, we examine the hypothesis that projects led by female task team leaders suffer a performance penalty relative to those led by males in environments with unequal gender norms favoring men, and that projects led by young TTLs suffer a performance penalty relative to those led by older TTLs in countries with age norms favoring seniority.

To test this hypothesis, we use a difference-in-differences design comparing how differences in project performance across TTLs vary with the alignment between their characteristics

and country-level norms. Specifically, we regress indicators of project performance, notably IEG outcome scores, on indicators for whether the TTL is female and whether the TTL is young, measures of gender inequality and gerontocratic norms in the recipient country, and, crucially, the interactions between these characteristics and the corresponding norms.

The primary empirical concern is that any observed performance differences reflect confounders rather than the interaction between leader characteristics and social norms. Differences in TTL ability, project type, or country characteristics could bias estimates. Of particular concern is differential selection: if the characteristics of TTLs vary systematically with the normative environment to which they are assigned, observed performance gaps may reflect selection rather than genuine differences. To explore this, we examine how proxies for TTL ability correlate with TTL characteristics (gender and age), country-level norms (gender inequality and gerontocracy), and their interactions, controlling for TTL, project, and country characteristics.

In our main analysis, we address potential selection bias by including extensive controls and fixed effects. Our most stringent specifications include both country and TTL fixed effects. Country fixed effects absorb differences in time-invariant country characteristics, such as geography or culture (to the extent it is time-invariant). TTL fixed effects remove bias from time-invariant characteristics of TTLs, such as innate ability, and ensure identification comes from comparing how the same TTL performs across countries with different levels of gender inequality.

Our preferred specification is:

$$\begin{aligned} Y_{ict} = \beta_1 \left(Female_i \times Gender\, Inequality_c\right) + \beta_2 \left(Young_i \times Gerontocracy_c\right) + \\ + \mu_i + c_c + \phi\, TTL_{it} + \theta\, P_{pt} + \psi_1\, C_{ct} + \psi_2\, \left(Female_i \times C_{ct}\right) + \psi_3\, \left(Young_i \times C_{ct}\right) + \varepsilon_{ict} \end{aligned}$$

Where Y_{ict} denotes the outcome for project p in country c initiated in year t supervised by TTL i; $Female_i$ is an indicator for a female TTL; $Gender\, Inequality_c$ is a time-averaged measure of gender inequality in the recipient country; $Young_i$ is an indicator for being less than 35 years old; $Gerontoracy_c$ is a measure ofattitudes towards the eldery in the recipient country, μ_i are TTL fixed effects and C_c are country fixed effects[3] TTL_{it} is a vector of time-varying TTL-level characteristics notably grade salary (which serves as a proxy of ability), dummies capturing prior project supervision experience (0, 1, 2 or more); P_{pt} is a vector of project-level characteristics notably log commitments, sector FE, and financing source (IDA, IBRD or both); and C_{ct} is a vector of time-varying country characteristics, notably GDP per capita, natural disaster hazard index, whether the country is at war, and whether it is in a debt or financial crises.

While this is a highly stringent specification, potential residual selection concerns remain. Assignment of TTLs to countries might reflect unobserved complementarities between leader characteristics and local environments. For example, female or younger TTLs operating in less favorable normative settings may be positively selected, such as particularly capable or motivated individuals, or negatively selected, such as those with fewer outside options.

To mitigate such concerns, we allow the returns to TTL characteristics to vary with

[3]Their inclusion precludes separate identification of TTL the direct effect of gender inequality on project performance.

other country attributes by including interactions between TTL characteristics and time varying country controls. This accounts for the possibility that differences in performance attributed to misalignment with social norms instead reflect differential effectiveness across environments. For instance, female TTLs may perform differently in fragile settings, or younger TTLs may be less effective in poorer or crisis affected countries. While TTL fixed effects address cross TTL selection, they may not fully capture TTL country specific complementarities.

To address these concerns, we implement an instrumental variables (IV) approach. Specifically, we instrument whether a project is assigned a female TTL using the gender composition of eligible TTLs by grade and by the unit hosting the project at the time of project approval. Intuitively, the instrument captures the likelihood that a project is led by a female TTL based on the pool of eligible female staff in that unit and grade, rather than project-specific characteristics. Because women have historically been less represented at senior grades (e.g., H-level lead specialists) than at junior professional grades (e.g., F-level professionals),[4] and because certain sectors, such as human development or education, employ more women than others, such as transport or infrastructure, the probability that a project is led by a female TTL varies both by grade and unit hosting the project.

The reasoning for young TTLs in gerontocratic countries follows the same logic described for female TTLs in gender unequal countries.

5 Who leads which projects?

To set the scene for the analysis that follows, we first examine the characteristics of female project managers and how they differ from men, as well as those of the project they manage. We subsequently assess whether women are less likely to lead projects in countries with unequal gender norms, and whether the ones that do have special characteristics that could impact project outcomes. This not only helps contextualize the results, but also helps us assess potential selection bias, which is one of the primary obstacles to identifying the impact of gender norms on gender differences in project performance.

5.1 Gender and age differences in Task Team Leaders and their Projects

Over the full sample period, only about one-third of World Bank projects have been led by women, despite women comprising XX percent of the institution's professional staff. Women are thus less likely to be project managers. Table 7 compares the characteristics of male and female task team leaders (TTLs) and the projects they manage. Panel A begins with demographic characteristics of the TTLs. On average, female TTLs are younger and have less experience, whether measured by years at the World Bank, professional grade, or number of projects managed. Figure 3 shows that women are more prevalent among lower-grade TTLs: 69% of entry-level analyst who are TTLs (GE level) are female, compared with 14% of lead specialists who serve as TTLs (GH level). This partly reflects the historic underrepresentation of women at higher grades, although gender representation

[4]See Das and Joubert (2023) for a comprehensive analysis.

has dramatically improved in more recent cohorts. Female TTLs also have somewhat lower educational attainment (despite most holding graduate degrees), earn lower salaries on average, and are less likely to be married.

Panel B documents differences in project characteristics. Importantly, we find no evidence of unconditional performance differences between male and female TTLs. Average project outcome ratings do not differ significantl across genders, and neither the expected nor the actual project duration varies by TTL gender. Projects led by female TTLs are somewhat smaller on average in terms of total amount of funding committed.

Gender sorting by sector is notable Women are more likely to lead in projects in social protection (48% of all TTL's are female) education (47%), and health (46%), and less likely to lead projects in agriculture (19%), transportation (16%) and energy and extractives (14%) (Appendix Figure 4). This sectoral sorting partly reflects the gender composition of staff working in those areas.

There is also strong sorting across countries, as is shown in panel C. Female TTLs tend to manage projects in countries with significantly lower gender inequality. These countries also have higher GDP per capita, stronger democratic institutions, lower incidence of conflict, and fewer financial crises. Across all four measures of recipient-country gender inequality, projects led by women are consistently in countries with lower inequality, though the magnitudes vary: ranging from 0.157 standard deviations lower according to the World Economic Forum (WEF) measure, 0.106 standard deviations lower according to the OECD's Social Institutions and Gender Index (SIGI), 0.08 lower according to the Women, Business and the Law (WBL) index, to 0.027 standard deviations lower according to the Gender Inequality Index (GII) published by the United Nations Development Programme (UNDP).

The likelihood that a project is led by a female TTL is strongly associated with the share of women among potential TTLs, proxied by the share of staff at the same grade level as the TTL in the unit hosting the project, as is shown in Panel D and Appendix Figure **??** which presents a binscatter of project leadership against the share of women at the relevant grade. We leverage this variation in the supply of potential TTLs in our instrumental variables strategy.

5.2 Which projects are led by female and young TTLs?

To assess which projects are led by women, Table 1 presents probit regressions in which the dependent variable equals one if the project manager is female. We add progressively more elaborate sets of explanatory variables. We are especially interested in examining whether women are less likely to lead projects in countries with unequal gender norms.

Column 1 only includes gender inequality using the gender gap index from the WEF as an explanatory variable. We focus primarily on this measure in the main text but present robustness to using alternative measures in the appendix. Gender inequality is negatively correlated with the probability of having a female TTL. A one standard deviation reduces the likelihood of having a female by 4.5% (the mean is 28.8%), but the predictive power of gender inequality appears limited given that the pseudo R2 is 0.00582.

The association between recipient country gender inequality and having a female TTL is robust to controlling for characteristics of the Task Team Leader, as is done in column 2, which furthermore shows that female TTLs are more prevalent at more junior professional

Table 1: Which projects are led by female or young TTLs? Probit models

	TTL is female			TTL is young		
	(1)	(2)	(3)	(4)	(5)	(6)
Gender Inequality (WEF)	-0.050***	-0.031**	-0.014		0.014*	
	(0.012)	(0.013)	(0.012)		(0.008)	
Gerontocracy		-0.008		-0.010	-0.008	-0.001
		(0.016)		(0.009)	(0.010)	(0.007)
Crisis		-0.003	-0.013		0.055***	0.053***
		(0.025)	(0.023)		(0.019)	(0.016)
IDA & IBRD		0.070	0.003		-0.096*	-0.025
		(0.043)	(0.040)		(0.057)	(0.042)
Has a Ph.D		-0.067*	-0.068**		-0.008	0.001
		(0.034)	(0.027)		(0.032)	(0.024)
%women in same grade in OUI			0.796***			
			(0.039)			
%young staff in same grade in OUI						0.533***
						(0.036)
MeanDep	0.248	0.263	0.263	0.200	0.235	0.236
N	1994	1769	1771	2065	1591	1727
R-squared	0.0415	0.192	0.350	0.0217	0.270	0.410
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Country controls		Yes	Yes		Yes	Yes
Project controls		Yes	Yes		Yes	Yes
Sector controls		Yes	Yes		Yes	Yes
Individual controls		Yes	Yes		Yes	Yes

Notes: Marginal effects are presented with standard errors clustered by staff member. Country controls include a dummy for financial crisis. Project controls include log commitments, financing source dummies, and sector. Individual controls include education, salary, grade fixed effects, sharing an official language and physical distance of the country of birth to the recipient country. Only country characteristics that are significant in at least one specification are shown. *, **, *** indicate significance at the 10%, 5%, and 1% significance level, respectively.

grades. The association attenuates but remains statistically significant when controlling for project characteristics, as is done in column 3. The likelihood of having a female TTL is lower when projects are larger and varies considerably across sectors. It is also robust to controlling for country characteristics as is done in column 4.

Finally, in column 5 we add controls for the share of women in the pool of potential TTLs by controlling for the share of women at the grade level of the TTL in the unit hosting the project, both itself and interacted with gender inequality. These variables, which will serve as our instrument, have very strong predictive power.

In summary, the negative association between gender inequality and the probability of a female TTL is robust across specifications.

5.3 Is there sorting of female and young Task Team Leaders (TTLs) across countries?

The previous subsection showed evidence of sorting by gender. Such sorting may create selection bias: female (young) Task Team Leaders (TTLs) who take on projects in gender-unequal (gerontocratic) countries may systematically differ from those working in more gender-equal (less gerontocratic) settings in ways that also affect project performance. For our purposes, the key concern is *differential* selection bias by gender: if there is sorting, then observed gender gaps in project performance could reflect selection rather than genuine performance differences.

To assess whether such selection bias might be at play, and whether it differs by gender or age, Table 2 examines how various proxies for TTL ability correlate with the TTL features, recipient country norms and the interaction between the two. Evidence of systematic correlations would indicate that our main estimates may be biased by differential sorting, whereas the absence of such patterns would mitigate concerns about confounding due to selection.

Overall, Table 2 clearly indicates no evidence of sorting based on the interaction between TTL features and cultural norms: female (young) project leaders in more gender-unequal (gerontocratic) environments seem neither positively nor negatively selected.

6 Do Cultural Norms Constrain Development Effectiveness?

This section first presents our main findings, then discusses robustness tests and end with heterogeneity analysis.

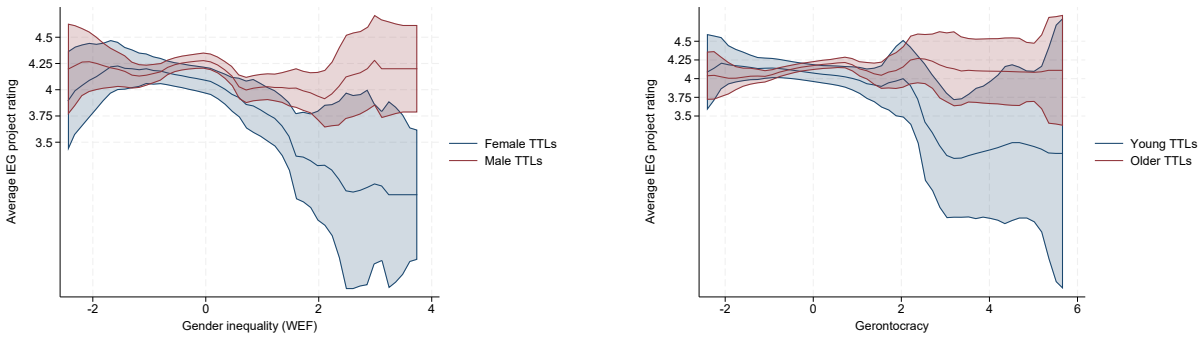
6.1 Main findings

Our main findings are presented in Table 3, which examines the determinants of IEG project outcome ratings using only projects for which we observe a single supervisor TTL. We first present OLS estimates and then discuss instrumental variables estimates. The first specification includes only approval-year dummies and a dummy for whether the TTL is a woman. The estimated female performance premium is small (0.058) and statistically

Table 2: Testing for Potential Ability Bias

	Salary	SRI (prior)	# Projects (prior)	IEG (prior)	PhD	Rookie
	(1)	(2)	(3)	(4)	(5)	(6)
Female TTL × Gender Inequality	228.6	0.0552	-0.0801	0.218	0.0644*	0.0195
	(960.4)	(0.0358)	(0.0590)	(0.206)	(0.0301)	(0.0289)
Female TTL	-1292.9	0.0123	0.265**	-0.0787	-0.0541	-0.0798**
	(1168.2)	(0.0567)	(0.0864)	(0.144)	(0.0516)	(0.0282)
Young TTL × Gerontocracy	176.1	-0.00168	-0.00830	-0.0956	0.00617	-0.0453
	(942.5)	(0.0467)	(0.0482)	(0.0953)	(0.0307)	(0.0290)
Young TTL	-3653.3**	0.0313	-0.296***	-0.696**	-0.00110	0.269***
	(1255.2)	(0.104)	(0.0594)	(0.205)	(0.0509)	(0.0352)
Mean dep. var.	45171.6	3.374	0.541	4.031	0.245	0.257
Observations	1941	532	1973	519	1973	1973
R^2	0.559	0.316	0.328	0.374	0.160	0.253

Notes: Each column reports a separate regression. All specifications include recipient country, sector, and approval year fixed effects, log commitments, war, financial crisis and financing source. Standard errors are clustered by TTL and recipient country. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.



(a) Female TTL and gender inequality **(b) Young TTL and gerontocracy**

Figure 1: IEG ratings, TTL features and local norms

Notes: The figures show local polynomial regressions of IEG ratings on recipient country norms by type of TTL. Panel a) shows the local polynomial regression of IEG ratings on gender inequality by gender of the TTL. Panel b) shows the local polynomial regression of IEG ratings on gerontocracy by age of the TTL.

insignificant (see also Table 7), indicating that, unconditionally, projects led by women do not perform significantly better or worse than those led by men.

Once we add TTL characteristics, project controls, and country fixed effects (column 2), the female performance penalty becomes larger and precisely estimated. All else equal, projects led by women receive outcome ratings that are 0.085 points lower than those led by men. This difference is modest relative to the dependent variable's standard deviation of 1.114 but not negligible. The contrast between the unconditional and conditional estimates suggests that women are more often assigned to projects that, on average, are easier to implement successfully, which may partly reflect concentration in more favorable country contexts and can mask underlying differences when no controls are included.[5]

Column 3 investigates a potential source of gender differences in performance, namely local gender norms. Both gender inequality itself and its interaction with being a female TTL are strongly statistically significant. A one standard deviation increase in gender inequality is associated with a 0.099-point lower outcome rating for projects led by men. For female TTLs, the reduction is even larger, with an additional 0.088-point decline on average. Thus, female-led projects perform worse than male-led projects in countries where gender norms favor men.

To assess the robustness of this key finding, column 4 adds controls for TTL and project characteristics. This reduces the coefficient on gender inequality, although it becomes more imprecisely estimated. At the same time, the coefficient on the interaction between female TTL and gender inequality becomes even more negative, reaching -0.107, and remains strongly significant. Column 5 introduces country fixed effects. This specification helps reduce concerns that other time-invariant country characteristics, such as the quality of long-run institutions or geography, could be driving the effect. The coefficient on the interaction between gender inequality and female TTL becomes slightly more negative and remains statistically significant. In other words, the finding that project led by women fare worse than those led by men in contexts where norms favor men seems robust.

The final OLS specification, presented in column 6, includes both TTL and recipient country fixed effects. This is a highly demanding specification, as identification is achieved by comparing performance differences within a given TTL across countries. Inclusion of TTL fixed effects reduces degrees of freedom considerably, but the coefficient on the interaction between gender inequality and female TTL becomes even more negative and remains statistically significant at the 10 percent level. A one standard deviation increase in gender inequality is associated with a 0.209 lower project outcome rating for female TTLs relative to male TTLs, on average. For example, a female TTL leading a project in Togo (90th percentile, index 1.38) would likely achieve a project outcome rating 0.46 points lower relative to her male counterparts than if she led the same project in Estonia (10th percentile of the gender inequality index, –0.82). Thus, female-led projects underperform relative to male-led projects in countries where local gender norms favor men, highlighting how cultural context can shape project effectiveness.

While our OLS specifications controlling for country and TTL fixed effects are already highly demanding, columns 7 and 8 present instrumental variables (IV) estimates of the same specifications as columns 5 and 6, respectively. Both IV estimates are considerably

[5]Another possibility is that they work on projects that are easier to implement. In the Appendix, we show that the female performance penalty arises only once we control for country fixed effects.

Table 3: Determinants of Project Outcomes

	IEG Outcome Rating					
	(1)	(2)	(3)	(4)	(5)	(6)
Female TTL × Gender Inequality	-0.134*		-0.136**	-0.122*	-0.144**	-0.397***
	(0.070)		(0.067)	(0.070)	(0.066)	(0.095)
Female TTL	-0.183***		-0.182***	-0.598	-0.262	
	(0.065)		(0.066)	(0.552)	(0.555)	
Gender Inequality	-0.060*		-0.056	-0.061		
	(0.034)		(0.035)	(0.038)		
Young TTL × Gerontocracy		-0.203***	-0.167***	-0.206***	-0.155**	-0.187**
		(0.046)	(0.054)	(0.065)	(0.060)	(0.077)
Young TTL		-0.090	-0.046	0.210	-0.320	-0.418
		(0.065)	(0.067)	(0.604)	(0.570)	(0.710)
Gerontocracy		0.032	0.015	0.052		
		(0.030)	(0.033)	(0.034)		
TTL, Project and country controls (cc)			Yes	Yes	Yes	Yes
Cc × young, cc × female				Yes	Yes	Yes
Country FE					Yes	Yes
TTL FE						Yes
mean Dep. Var.	4.140	4.121	4.142	4.145	4.144	4.145
Observations	1,994	2,132	1,992	1,946	1,941	1,927
N. TTL	524	530	524	521	521	507
N. rec. countries	109	129	108	107	102	102
R-squared	0.045	0.036	0.048	0.094	0.216	0.521

Notes: Standard errors clustered two ways by TTL and recipient country are presented in parentheses. All regressions include approval year fixed effects. Project controls include financing source dummies, sector dummies and log commitments. TTL controls include salary, a dummy for having a PhD and grade FE. Country controls include being at war, log GDP per capita, financial crisis and their interactions with young TTL and female TTL.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

more negative and strongly statistically significant. The Kleibergen–Paap statistics indicate that the instrument is not weak. Without country fixed effects, a one standard deviation increase in gender inequality is associated with a 0.178-point lower project outcome rating for female TTLs, compared to 0.114 in the corresponding OLS specification. Controlling for TTL fixed effects, the effect rises to 0.424, relative to 0.209 in OLS. In other words, the OLS estimates are slightly biased downward. One possible explanation is that women who take on projects in gender-unequal countries are, on average, particularly adept at navigating the local context, which partially offsets the observed performance penalty.

6.2 IV estimates

We believe that the inclusion of country and TTL fixed effects (joint to the inclusion of a wide variety of controls) and, most importantly, the absence of evidence of sorting of female and young TTLs across countries based on ability, make our main specification reliable and convincing.

This said, it is admittedly hard to rule out all possible alternative explanations, no matter how contrived, based on our FE specification. Hence, to address these residual concerns, we deviced an IV strategy based on the amazing depth of information that the HR data provide us. Specifically, we instrument the gender and the age of the TTL based on the pool of candidate TTLs, thus bypassing concerns related to sorting within such pool. The pool of candidate TTLs is the set of officers located in the same unit (OUI) and grade of the TTL. We access the demographics of these officers and generate the share of female officers and the share of young officers to instrument whether the TTL is female and/or young.

Table 4 shows the results. Results associated with the IV strategy are broadly consistent with those obtained with the OLS strategy.

Table 4: IV estimates

	IEG Outcome Rating			
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
Female TTL × Gender Inequality	-0.147**	-0.249**	-0.407***	-0.545***
	(0.065)	(0.095)	(0.094)	(0.141)
Female TTL	-0.165**	-0.486**		
	(0.077)	(0.191)		
Young TTL	-0.026	0.116	-0.005	0.228
	(0.073)	(0.187)	(0.099)	(0.329)
Young TTL × Gerontocracy	-0.178***	-0.262***	-0.207***	-0.488***
	(0.055)	(0.083)	(0.071)	(0.093)
TTL FE			Yes	Yes
Observations	1,941	1,941	1,927	1,927
R-squared	0.212	-0.001	0.519	0.004
mean Dep. Var.	4.144	4.144	4.145	4.145
sd Dep. Var.	1.116	1.116	1.117	1.117
Number of TTLs	521	521	507	507
Recipient countries	102	102	102	102
KP stat	–	34.13	–	40.58

Notes: Standard errors clustered two ways by TTL and recipient country are presented in parentheses. All regressions include approval year fixed effects. Project controls include financing source dummies, sector dummies and log commitments. TTL controls include salary, a dummy for having a PhD and grade FE. Country controls include being at war, log GDP per capita, financial crisis and their interactions with young TTL and female TTL.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

One concern that people might have with this IV strategy is that units (OUIs) might specialize in projects located in certain parts of the world. Specifically, if some units work systematically with gender unequal countries, while others work systematically with gender equal ones, and officers know this in advance, they might sort themselves across units according to their preferences and, in turn, this sorting could somehow bias our estimates. Against this interpretation there is of course the fact our empirical strategy relies on TTLs who move across units and that anyway TTLs have been shown not to sort based on ability. This notwithstanding, in additional analysis, we control for the share of female and young officers in a given unit over the entire sample period and obtain results that are not only qualitatively but also quantitatively very similar.

6.3 Alternative indexes

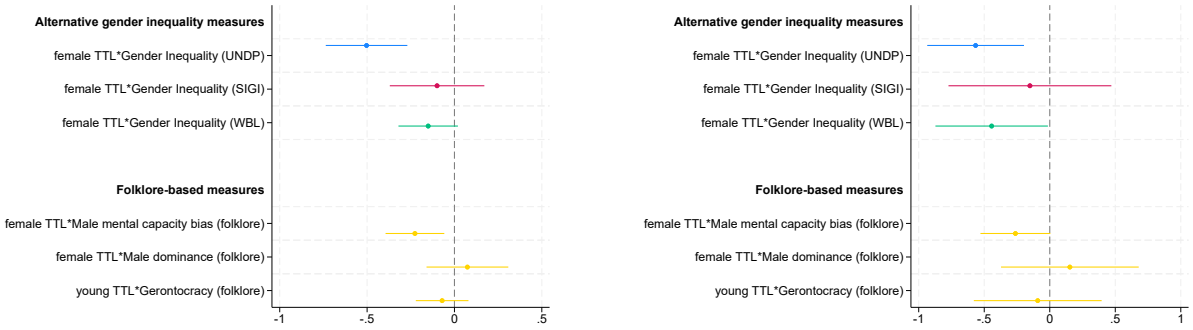
As anticipated in the Data Section, one may want to measure gender inequality and gerontocracy in other ways.
For gerontocracy, the alternatives are very limited. The only other index that covers many middle and low income countries is the Folklore-based index provided by Saenger (2023). The advantage of this index is that it is predetermined. The disadvantage is that, well, it is based on the assumption that norms common centuries ago are still strong enough nowadays as to influence current behavior.

For gender inequality, there are several alternative options: the Gender Inequality Index (GII) by the UNDP, the Social Institutions and Gender Index (SIGI) by the OECD, or the Women, Business and the Law (WBL) by the World Bank. In addition, we also use two Folklore-based gender inequality indexes provided by Michalopoulos and Xue (2021): the male mental capacity bias and the male dominance indexes. The former captures the share of motifs in which men are portrayed as more intelligent and less naive than women. The latter is the share of motifs where men are depicted as more dominant, less submissive, less engaged in domestic affairs, and more physically active than women.

Figure **??** shows the OLS and IV estimates associated with these alternative indexes.

For gender inequality, results are robust when using the UNDP and WBL index, but not when using the OECD index. Overall, this seems reassuring. In terms of Folklore-based indexes, results are robust when using the male mental capacity bias but not when using the male dominance index. This is consistent with the view that the interaction between TTL and implementing agency is based on intellectual engagement and persuasion (more often via electronic devices than in person) rather than based on physicality.

For gerontocracy, results based on the Folklore-based index are generally weaker.



(a) OLS estimates **(b) IV estimates**

Figure 2: Alternative measures of age and gender norms

Notes: Plot shows selected coefficients from regressions using alternative measures of age and gender norms. Each color denotes a separate regression. All specifications control for country and TTL fixed effects, as well as country, TTL and project characteristics. Horizontal lines denote 90% confidence intervals.

7 Other cultural attributes

In Table 5, we examine whether more subtle dimensions of cultural alignment between task team leaders (TTLs) and recipient countries are associated with project performance. Column (1) tests whether outcomes are higher when the TTL speaks an official language of the recipient country. Column (2) examines shared nationality between the TTL and the recipient country. Column (3) considers whether TTLs originating from countries with a history of bilateral conflict with the recipient country perform differently. Column (4) assesses whether cross-country trust may mediate effectiveness. Column (5) controls for the geographic distance between the TTL's country of origin and the recipient country. Columns (6) and (7) incorporate measures of cultural distance, based on the indices developed by Wacziarg and Culture Wizard, respectively.
Across specifications, none of these measures significantly predicts project performance.[6]

[6]The absence of an effect of shared nationality contrasts with the findings of Bau et al. A key distinction is that our specification includes project manager fixed effects. Omitting these fixed effects yields a negative association similar to the one they report, suggesting that failure to account for time-invariant manager heterogeneity may bias estimates.

Table 5: Other measures of cultural alignment

	IEG Outcome Rating (1–6)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female TTL × Gender Inequality	-0.395***	-0.382***	-0.399***	-0.380***	-0.411***	-0.365***	1.313	-0.401***	-0.474***
	(0.095)	(0.096)	(0.095)	(0.096)	(0.118)	(0.118)	(1.871)	(0.094)	(0.118)
Young TTL × Gerontocracy	-0.179**	-0.188**	-0.187**	-0.186**	-0.154	-0.132	-0.513	-0.177**	-0.156
	(0.082)	(0.077)	(0.077)	(0.078)	(0.184)	(0.149)	(0.305)	(0.083)	(0.176)
Young TTL	-0.521	-0.443	-0.432	-0.413	-0.270	-0.283	5.623**	-0.512	-0.464
	(0.711)	(0.705)	(0.709)	(0.718)	(0.931)	(0.846)	(2.756)	(0.712)	(0.896)
Same language	-0.281***							-0.328***	-0.465**
	(0.107)							(0.116)	(0.176)
Same nationality		-0.086						-0.026	-0.243
		(0.118)						(0.136)	(0.246)
Prior conflict			-0.203					-0.296*	-0.307
			(0.156)					(0.162)	(0.244)
Physical distance				-0.136				-0.459	-0.298
				(0.373)				(0.404)	(0.496)
Trust					0.247				0.459
					(0.793)				(0.869)
Cultural Distance (Wacziarg)						0.070			-0.119
						(0.106)			(0.124)
Cultural Distance CW							9.411**		
							(4.239)		
TTL, project and country controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country controls × TTL characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TTL FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Approval year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
mean Dep. Var.	4.143	4.143	4.145	4.143	4.236	4.203	4.007	4.143	4.236
Observations	1,915	1,915	1,927	1,915	1,303	1,445	147	1,915	1,303
N. TTL	503	503	507	503	372	400	43	503	372
N. recipient countries	101	101	102	101	61	69	37	101	61
R-squared	0.522	0.520	0.522	0.520	0.536	0.525	0.917	0.524	0.542

Notes: Standard errors clustered two ways by TTL and recipient country are presented in parentheses. All regressions include approval year fixed effects. Project controls include financing source dummies, sector dummies and log commitments. TTL controls include salary, a dummy for having a PhD and grade FE. Country controls include being at war, log GDP per capita, financial crisis and their interactions with young TTL and female TTL.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

8 How costly are cultural constraints for recipient countries?

If we assume that the money spent on projects that fail (i.e. which have an IEG outcome rating score of 3 or lower) is wasted entirely, we can calculate the costs of gender inequality and age discrimination in terms of wasted development spending.
First, we collapse project ratings into a binary indicator: success vs failure. According to this metric, the actual success rate: 73.9 percent. We then ask ourselves what would happen if all countries had the gender inequality and age norms set at the 10th percentile. If that was the case, our results indicate that the success rate would be 77.2 percent. In other words, the share of projects that fail due to gender inequality is 3.2 percent. The "cultural cost" can be decomposed into a gender component (2.8 percent) and an age component (0.5 percent).
We can then use project commitments to convert this percentage of "missed successes" into a monetary counterpart of 1.38 percentage points, which, in our sample, corresponds to 2.4 billion USD.

9 How costly are cultural constraints for managers?

We examine how cultural misalignment affects the career progression of project leaders, focusing on performance ratings and promotions. World Bank staff receive annual SRI ratings from 1 to 5, with ratings of 4 or 5 indicating exceptional performance and determining salary increases. Table 6 analyzes whether cultural misalignment influences the likelihood that TTLs receive a high rating during the project and at closing, when project success is assessed. We study ratings in the closing year and in the subsequent year, when evaluation results become known. All regressions include controls for TTL and country characteristics, as well as TTL and country fixed effects, so that estimated differences reflect within-TTL variation over time.
The results show that women are significantly less likely to receive a top performance rating when leading projects in more gender-unequal countries. A one-standard-deviation increase in gender inequality reduces the probability of receiving a high rating by XXX% in the year the project closes, consistent with these projects being less successful. Women do not receive significantly different ratings during the implementation period, when project outcomes are not yet known, and the effects on promotions are not statistically significant. By contrast, young TTLs do not experience a comparable penalty. These findings suggest that exposure to hostile cultural norms may contribute to the gender pay gap in international organizations and slow women's career advancement.

Table 6: Cultural mismatch and career progression

	During		At Closing		Closing + 1		Closing + 2	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: officer's yearly evaluation (SRI rating, continuous)								
Female TTL × Gender Inequality	-0.043	-0.031	-0.111	-0.339**	-0.166**	-0.142	0.035	0.030
	(0.028)	(0.048)	(0.067)	(0.148)	(0.071)	(0.152)	(0.065)	(0.102)
Young TTL × Gerontocracy	0.033	0.056	0.041	0.044	0.053	-0.021	-0.006	-0.016
	(0.024)	(0.046)	(0.040)	(0.072)	(0.046)	(0.094)	(0.039)	(0.064)
Young TTL	0.059	0.178	0.135	-0.066	-0.058	-0.247	-0.051	-0.056
	(0.041)	(0.117)	(0.082)	(0.258)	(0.084)	(0.216)	(0.082)	(0.214)
mean Dep. Var.	3.459	3.459	3.536	3.536	3.534	3.534	3.520	3.520
Panel B: officer's yearly evaluation (SRI rating, binary)								
Female TTL × Gender Inequality	0.010	0.002	-0.123**	-0.363***	-0.097*	-0.097	0.027	0.002
	(0.028)	(0.064)	(0.055)	(0.115)	(0.055)	(0.113)	(0.044)	(0.069)
Young TTL × Gerontocracy	-0.002	-0.037	0.072**	0.080	0.030	-0.013	-0.028	-0.026
	(0.019)	(0.044)	(0.036)	(0.058)	(0.035)	(0.058)	(0.027)	(0.046)
Young TTL	0.016	-0.021	0.086	-0.065	-0.080	-0.272*	-0.132**	-0.245
	(0.035)	(0.109)	(0.062)	(0.186)	(0.061)	(0.151)	(0.060)	(0.158)
mean Dep. Var.	0.152	0.152	0.432	0.432	0.429	0.429	0.426	0.426
Observations	1,467	1,467	1,353	1,353	1,409	1,409	1,370	1,370
Approval year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TTL controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Project Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TTL FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N. TTL	423	423	405	405	424	424	411	411
N. recipient countries	102	102	101	101	100	100	100	100
Observations	1,915	1,997	1,825	2,073	1,943	2,015	1,855	2,098
R-squared	0.576	0.580	0.582	0.577	0.589	0.595	0.590	0.586
TTL FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Recipient countries	104	114	100	121	105	116	101	123

Notes: Standard errors clustered two ways by TTL and recipient country are presented in parentheses. All regressions include approval year fixed effects. Project controls include financing source dummies, sector dummies and log commitments. TTL controls include salary, a dummy for having a PhD and grade FE. Country controls include being at war, log GDP per capita, financial crisis and their interactions with young TTL and female TTL.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

10 Conclusion

This paper shows that cultural misalignment between development professionals and the environments in which they operate constrains aid effectiveness. Leveraging the rotation of World Bank Task Team Leaders across countries, we compare the performance of the same individuals working under similar incentives in different cultural contexts.

Using these data, we first show that assignment patterns in part reflect local norms: women are less likely to lead projects in more gender-unequal countries, with no comparable sorting by age depending on views of the elderly. These differences do not reflect ability.

Second, cultural misalignment reduces performance. Projects led by female TTLs perform worse in more gender-unequal environments, and projects led by younger TTLs perform worse in societies that favor seniority. We find no analogous effects for broader measures of cultural distance

Our results do not imply that donors should avoid appointing women to lead projects in countries where norms favor men, as such appointments may generate demonstration effects and other positive spillovers beyond the scope of this study.[7] Rather, they show that entrenched gender norms can undermine both equity and efficiency. The success of development interventions depends, in part, on addressing these social barriers.

[7]Exposing local implementation agencies to female leaders could change beliefs in female's ability (Beaman et al 2009), parents' aspirations (Beaman et al 2012) and, in turn, local gender norms.

References

Andersen, Jørgen Juel, Niels Johannesen, and Bob Rijkers, "Elite capture of foreign aid: Evidence from offshore bank accounts," *Journal of Political Economy*, 2022, *130* (2), 388–425.

Arndt, Channing, Sam Jones, and Finn Tarp, "What is the aggregate economic rate of return to foreign aid?," *The World Bank Economic Review*, 2016, *30* (3), 446–474.

Ashraf, Nava, Oriana Bandiera, Virginia Minni, and Víctor Quintas-Martínez, "Gender gaps across the spectrum of development: Local talent and firm productivity," 2024.

Ashton, Louise, Jed Friedman, Diana Goldemberg, Mustafa Zakir Hussain, Thomas Kenyon, Akib Khan, and Mo Zhou, "A Puzzle with Missing Pieces: Explaining the Effectiveness of World Bank Development Projects," *The World Bank Research Observer*, 2023, *38* (1), 115–146.

Bank, World, "Women, Business and the Law 2023," 2023. Accessed: 2025-08-04.

Bau, Natalie, Sara Lowes, and Eduardo Montero, "Culture, Policy, and Economic Development," Technical Report, National Bureau of Economic Research 2025.

Burnside, Craig and David Dollar, "Aid, policies, and growth," *American economic review*, 2000, *90* (4), 847–868.

_ **and** _ , "Aid, policies, and growth: reply," *American economic review*, 2004, *94* (3), 781–784.

Centre, OECD Development, "Social Institutions and Gender Index 2023 Global Report," 2023. Accessed: 2025-08-04.

Chiplunkar, Gaurav and Pinelopi Koujianou Goldberg, "Aggregate implications of barriers to female entrepreneurship," *Econometrica*, 2024, *92* (6), 1801–1835.

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

Ek, Andreas, "Cultural values and productivity," *Journal of Political Economy*, 2024, *132* (1), 295–335.

Forum, World Economic, "Global Gender Gap Report 2021," 2021. Accessed: 2025-08-04.

Geli, Patricia, Aart C. Kraay, and Hoveida Nobakht, "Predicting World Bank Project Outcome Ratings," Technical Report 7001, The World Bank 2014.

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow, "The allocation of talent and us economic growth," *Econometrica*, 2019, *87* (5), 1439–1474.

Kilby, Christopher, "Assessing the impact of World Bank preparation on project outcomes," *Journal of Development Economics*, 2015, *115*, 111–123.

Klasen, Stephan, "The impact of gender inequality on economic performance in developing countries," *Annual Review of Resource Economics*, 2018, *10* (1), 279–298.

Lansing, J Stephen, *Priests and programmers: technologies of power in the engineered landscape of Bali*, Princeton University Press, 2009.

Lowes, Sara and Eduardo Montero, "The legacy of colonial medicine in Central Africa," *American Economic Review*, 2021, *111* (4), 1284–1314.

McGuirk, Eoin F and Nathan Nunn, "Development mismatch? Evidence from agricultural projects in pastoral Africa," Technical Report, National Bureau of Economic Research 2024.

Michalopoulos, Stelios and Melanie Meng Xue, "Folklore," *The quarterly journal of economics*, 2021, *136* (4), 1993–2046.

Moscona, Jacob, "The management of aid and conflict in Africa," *American Economic Journal: Economic Policy*, 2025, *17* (4), 228–259.

_ **and Awa Ambra Seck**, "Age set versus kin: Culture and financial ties in East Africa," *American Economic Review*, 2024, *114* (9), 2748–2791.

Mulligan, Casey B and Xavier Sala i Martin, "Gerontocracy, retirement, and social security," 1999.

Nguyen, Kieu-Trang, "Trust and innovation within the firm: Evidence from matched CEO-Firm data," Technical Report, Working paper. London School of Economics 2018.

Programme, United Nations Development, "Human Development Report 2020: The Next Frontier—Human Development and the Anthropocene," 2020. Accessed: 2025-08-04.

Qian, Nancy, "Making progress on foreign aid," *Annu. Rev. Econ.*, 2015, *7* (1), 277–308.

Rossignol, Etienne Le, Sara Lowes, and Eduardo Montero, "Fallow lengths and the structure of property rights," Technical Report, National Bureau of Economic Research 2024.

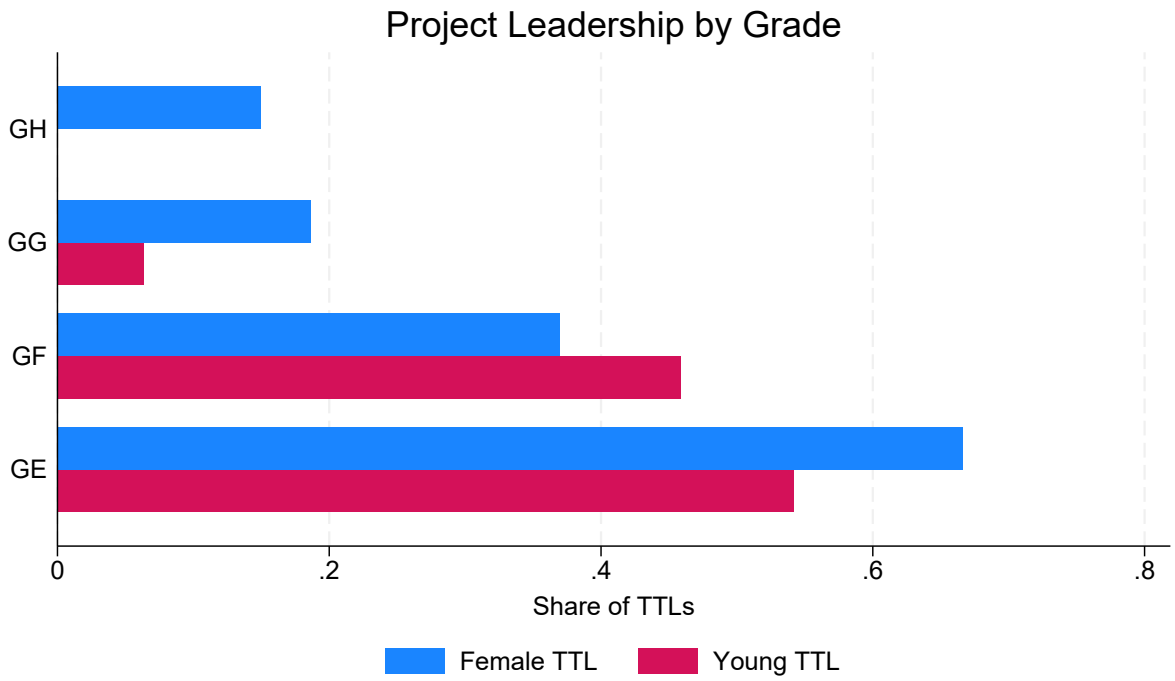
Saenger, Leo Yamamoto, "Respect Your Elders? The Economic Origins and Political Consequences of Attitudes Toward the Aged," 2023.

11 Appendix

Table 7: Comparison of Characteristics by Gender

	Men	Women	Difference	p-value
Age	43.216	40.467	-2.749	0.000
PhD	0.249	0.216	-0.033	0.454
Masters	0.516	0.659	0.143	0.005
Is married	0.843	0.509	-0.334	0.000
Salary	47490.449	40026.390	-7464.060	0.000
1 prior project	0.180	0.214	0.034	0.187
2 prior projects	0.098	0.090	-0.007	0.747
gen - analyst	0.020	0.118	0.098	0.000
GF - specialist	0.220	0.384	0.164	0.000
GG - senior specialist	0.655	0.446	-0.208	0.000
GH - expert	0.078	0.041	-0.037	0.124
IEG Rating (1-6 from worst to best)	4.148	4.017	-0.131	0.057
Project Delay (days)	580.753	544.219	-36.534	0.299
Project Delay >0	0.761	0.743	-0.018	0.472
Agriculture	0.161	0.074	-0.087	0.000
Education	0.081	0.175	0.095	0.004
Energy & Extractives	0.126	0.055	-0.071	0.002
Financial Sector	0.029	0.055	0.027	0.058
Health	0.047	0.161	0.114	0.000
Industry Trade Services	0.043	0.072	0.028	0.100
Info & Communication	0.011	0.020	0.009	0.340
Public Admininstration	0.136	0.140	0.004	0.850
Social Protection	0.042	0.094	0.052	0.027
Transportation	0.187	0.079	-0.108	0.000
Water Sanitation Waste	0.137	0.074	-0.063	0.003
IDA	0.477	0.450	-0.027	0.492
IBRD	0.491	0.506	0.015	0.718
IDA & IBRD	0.032	0.044	0.013	0.331
Log Commitments (USD)	17.608	17.352	-0.256	0.006
GDP per capita (constant 2015 USD)	3.032	3.347	0.315	0.254
CPIA	3.398	3.464	0.066	0.279
Polity	1.685	3.024	1.339	0.013
War	0.018	0.006	-0.012	0.021
Crisis	0.502	0.506	0.004	0.896
Natural Hazards	5.097	5.073	-0.024	0.873
Gender Inequality (WEF)	0.006	-0.248	-0.254	0.001
Gender Inequality (SIGI)	0.004	-0.086	-0.090	0.378
Gender Inequality (UNDP)	-0.052	-0.198	-0.146	0.083
Gender Inequality (WBL)	0.011	-0.114	-0.103	0.204

Figure 3: Distribution of female and young TTLs by grade



11.1 Implicit Association Test (IAT)

Figure 4: IAT-based Gerontocracy: distribution

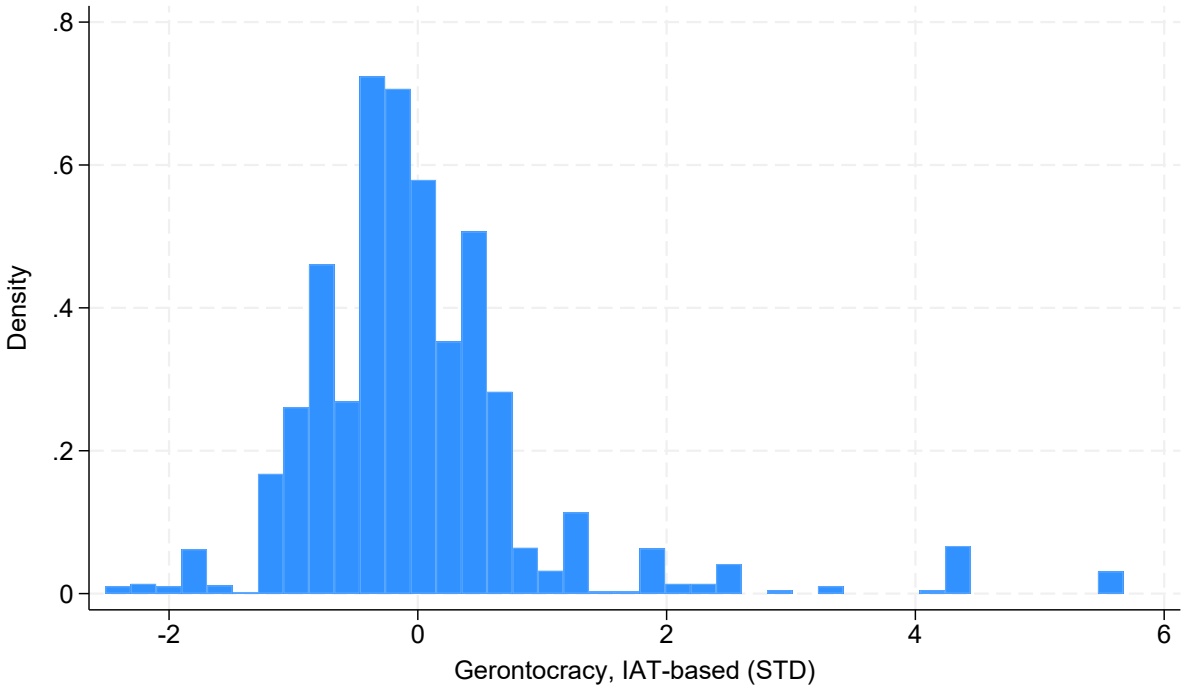


Figure 5: IAT-based Gerontocracy vs alternative measures

