The Power of Locality: How Local Staff Enhance Performance in International Organizations

Paper submitted to the PEIO conference 2025

Steffen Eckhard*, Sebastian Korb*, Katelyn Cassin[+] & Daniel Baumann*

**Zeppelin University Friedrichshafen*

[+]*University of Ottawa*

Abstract: Around two thirds of the staff body of international organizations (IO) are locally recruited. As citizens of the host state, there have been questions about their impartiality vis-à-vis the IO mandate and, ultimately, about their impact on the performance of IO policy implementation. This paper presents first conclusive evidence on the positive impact of local staff on project performance, drawn from a mixed method research design of IOs in the United Nations (UN) system. First, a quantitative analysis that combines novel data on the staffing mix in 998 unique office locations in 185 countries across five IOs and 16 years with performance measures drawn from 5,176 IO project evaluations reveals a significant positive relationship. Second, a case study of two key IOs in Lebanon presents qualitative validation of our quantitative findings and the expected causal mechanisms. Together, these findings provide robust evidence, suggesting that local staff enhance operational effectiveness through two mechanisms: provision of contextual knowledge and facilitation of relations with host communities. These findings underscore the importance of leveraging local human resources to optimize international policy implementation efforts in diverse geopolitical landscapes.

Words: 11798

Keywords: International organizations, Local staff, Project performance, Natural language processing

Introduction

Many international organizations (IOs) are heavily focused on field-level activities, which encompass numerous programs and projects across various countries and thematic areas. When conducting such "'boots-on-the-ground' implementation work" (Parizek 2024: 3), these operational IOs face high demands for country- and context-specific knowledge, a crucial source being their employees. As international staff frequently rotate, locally hired employees become the core of context-specific knowledge and institutional memory. Some scholars argue that this dynamic can create informational discrepancies, leading to principal-agent problems and rogue agent behavior. Known as "local capture," this phenomenon suggests that unrestricted local staff might negatively impact IO performance (Dávid-Barrett and Fazekas 2020; Eckhard 2014; Heinzel 2022a; Takasaki 2011; Winters 2014). But there are also many case studies and some preliminary quantitative findings highlighting the beneficial effects local staff have on project implementation (Campbell 2018; Coleman 2024; Eckhard 2019; Heinzel 2022b; Honig 2018; Sundberg 2023).

In this paper, we argue that the benefits of local staff outweigh the risks they pose to IO performance. While we are not the first to advocate for local hiring and the importance of local expertise (Campbell 2018; Coleman 2024; Eckhard 2019; Sundberg 2023), our study addresses limitations faced by previous research due to data availability. First, project performance data is typically available for only a few International Development Banks which publish standardized project performance ratings based on external ex-post evaluations (Heinzel 2022a; Honig 2018, 2019). Second, project-specific staff data is generally unavailable. Some scholars have addressed this gap by focusing, in the case of the World Bank, on the country experience of the heads of individual projects (Heinzel 2022b). Others have integrated the role of local knowledge into theoretical frameworks, using proxies like IO flexibility to discuss country-specific expertise (Honig 2018; Marchesi and Masi 2021). Overall, this means that we are still missing conclusive evidence, drawn from a diverse sample of IOs and country conditions, on the performance impact of the local staff.

In addressing this gap, we present findings from a mixed-method analysis that combines qualitative and quantitative insights on our central research question, how the presence of local staff in IO country offices affects the performance of their operations. First, we conduct a quantitative analysis employing natural language processing (NLP) combined with a novel dataset on United Nations (UN) system human resources. We use deep learning-based text classification to measure IO project performance for a larger number of IOs. Like the existing

studies, our measure draws on ex-post evaluations conducted by external evaluators. But we use NLP to extract performance scores from the qualitative data in evaluation reports rather than employing a human-provided performance rating which are only available for the World Bank and a few others. Next, we draw on UN system staff statistics for five IOs and across 16 years. We measure the staffing mix at 827 unique office locations in 172 countries and match them with 5,176 IO project evaluations. Employing multilevel and fixed effects regression models, we provide robust evidence suggesting a positive impact of higher shares of local staff on project performance. Contrary to expectations, however, the positive impact of local staff does not increase linearly with greater country fragility; instead, it decreases in the most fragile countries.

Second, these quantitative insights are complemented by a qualitative case study. We selected Lebanon as a 'typical' country case for IO policy implementation, focusing on the activities of IOs in the fields of peacekeeping and development assistance. The findings show that local staff enhance IO effectiveness through two main mechanisms: contextual knowledge, and relation-building with host communities. Their deep cultural and historical understanding enables them to navigate complex socio-political environments effectively, facilitating smoother project implementation and fostering stronger relationships with host communities. The case study therefore illustrates the causal relationship between local staff and project success and validates our quantitative findings.

Our approach adds to the current debate by including relatively understudied IOs and introducing a more precise measure of local staff based on actual staff composition at the level of individual offices. The combination of qualitative and quantitative analysis offers robust evidence on the causal relationship between the presence of IO staffers in IO country operations and the performance of these entities in achieving their goals.

Our findings are also of significant practical relevance for IOs in the UN system, which have been hiring international staffer at a higher rate than their local peers between 2007 and 2022. In that period, the total staff of the UN system grew from just over 80,000 to around 125,000 while the relative share of local employees dropped from 69% to 63%. This trend appears prevalent across the UN system and is evident in large IOs from various sectors (Appendix I). With out insights, we recommend UN system IOs to critically reflect on their staffing practices, to carefully evaluate which job profiles require international or locally recruited candidates.

This paper is organized as follows: The next section, we provide an overview of the literature on IO performance and staff characteristics. We then develop our theoretical argument based

on the literature concerning local staff in general, and particularly in fragile contexts. Following this, the quantitative analysis of five UN system IO operations tests our theoretical expectations. The findings and the causal mechanisms are further validated by the qualitative case study of two IOs in Lebanon. After discussing the results and limitations, the final section discusses theoretical and practical implications.

IO Performance and Staff Characteristics

Researchers commonly define the performance of a public organization as its capacity to achieve predetermined objectives (Gutner and Thompson 2010; Heinrich 2012). According to Lipson, it concerns "an organization's use of its resources, technology, and relationships with its organizational environment to achieve collective goals" (2010: 256). Conceptually, IO performance comprises multiple dimensions, including its throughputs (i.e. efficiency and effectiveness of internal processes), outputs (i.e. short-term products or services that an organization generates), and outcomes (i.e. external results of the organization's functioning, such as the medium- and long-term social, economic, or environmental impacts). In this paper, we focus on the outcome dimension of IO performance. Empirical studies have revealed significant variation in the performance of IOs (Honig 2019; Panke 2021; Tallberg et al. 2016), raising questions about the underlying causes of these disparities.

Traditional cross-country outcome studies, such as those emerging from the aid effectiveness debate spanning the 1980s to the 2000s, primarily concentrated on the effects of country characteristics on the annual impact of international development aid flows (Collier and Dollar 2004; Dollar and Pritchett 1998; Hansen and Tarp 2000; Mosley 1986). In contrast, contemporary scholarship has redirected its focus towards the administrations, policies, and activities of IOs themselves. Gulrajani aptly describes this shift as the "bureaucratic turn" (2017: 375). The increase of project evaluations at IOs, particularly those working in development, spurred a growing body of literature using evaluation ratings as outcome measure, specifically since the World Bank (WB) began to publish their evaluation ratings together with a vast number of project level variables. Besides organizational characteristics like funding patterns (Heinzel, Cormier, and Reinsberg 2023), transparency mechanisms (Honig, Lall, and Parks 2023) and levels of autonomy (Honig 2018; Lall 2017; Sommerer et al. 2022), numerous project-level factors have since been identified to affect outcome ratings. These include quality of supervision (Buntaine and Parks 2013; Ika 2015), monitoring system (Raimondo 2016), number of co-founders (Winters 2019), preparation time (Kilby 2015), type of implementing

partner (Shin, Kim, and Sohn 2017), and manager qualification (Heinzel and Liese 2021), early warning indicators (Denizer, Kaufmann, and Kraay 2013), project design (Ashton et al. 2023; Assefa, Rivera, and Vencatachellum 2014), duration (Bulman, Kolkma, and Kraay 2017) and project size (Honig 2020).

An early finding of this research is that country context variables affect outcome ratings far less than project management-related variables (Ashton et al. 2023; Denizer, Kaufmann, and Kraay 2013). These findings seem to hold for both WB and Asian Development Bank (Bulman, Kolkma, and Kraay 2017).

Different studies address the effects of staff characteristics on IO performance (von Billerbeck et al. 2024; Denizer, Kaufmann, and Kraay 2013; Eckhard 2014, 2019; Heinzel 2022b; Heinzel and Liese 2021). Heinzel (2022b) finds that the country experience of WB Task Team Leaders (project heads) also contributed to better outcome performance. On the organizational level, Lall (2017) demonstrates that more context specific staff qualification indicates higher de facto autonomy of IO administrations, which in turn improves IO performance. In line with these findings, Honig (2020) finds the presence of higher-ranking staff in the implementation country, especially during project preparation, to improve outcome performance of WB projects. Heinzel and Liese (2021) conclude that WB TTL qualifications enhance the performance of the implementation partner. Overall, staff characteristics account for about the same amount of project-level performance variance as all country-level characteristics combined (Ashton et al. 2023; Bulman, Kolkma, and Kraay 2017; Denizer, Kaufmann, and Kraay 2013).

But while the mentioned studies investigate staff and their qualifications often based on WB directorial staff (Country Directors or TTLs) (Heinzel 2022b; Heinzel and Liese 2021; Honig 2020; Reinsberg et al. 2024), there are no comparative investigations on the effect of staff composition in terms of staff citizenship on IO performance. This is all more surprising as numerous studies on IO staffing and representation (Eckhard 2014; Eckhard and Steinebach 2021; Hong 2017; Parizek and Stephen 2021, 2021; Weaver et al. 2022) highlight the relevance of national staff presence in IO administrations and projects.

Potential Benefits and Downsides of National Staff

The recruiting of host country citizens into international bureaucracies is a long ongoing practice (Parizek 2017; Parizek and Stephen 2021; Sundberg 2023). In the UN system the staff category of the National Professional Officer (NPO) was introduced in the 1960s and further formalized in the 1980s (Kullenberg 2016). NPOs undertake substantive roles, including

positions such as political affairs officer, legal officer, medical officer, or humanitarian affairs officer, typically under the supervision of International Professional (IP) staff. Another locally recruited staff group is the General Services (GS). GS staff handle technical or support functions, encompassing administrative, secretarial, and clerical duties. At higher pay grades, GS staff also engage in substantive tasks, such as program assistance, research, and community liaison roles (Eckhard 2021). Locally recruited staff constitutes the majority in UN field stations and made up over 63% of all UN systems employees in 2022 (UNCEB 2023). But while the share of NPOs and GS within the UN system is substantial, there is still strong variance between IOs and offices (Eckhard and Steinebach 2021). How this variance affects the performance of the projects managed by field offices has till now not been investigated.

Some scholars discussed the risk of local capture and biased decision-making by local bureaucrats as potentially harmful to IO performance (Eckhard 2014; Takasaki 2011; Winters 2014). Local capture can take the form of misappropriating funds through procurement processes (Winters 2014). For example, suppliers can inflate contract values or deliver lower-quality products (Olken 2007). There is a common concern that capture is more likely if national suppliers are involved because of their potential ties to recipient government officials (Dávid-Barrett and Fazekas 2020; Heinzel 2022a). To prevent local capture IOs install safeguards of international staff, such as frequent rotation from duty stations[1], attractive financial compensations[2] or limitations in discretion and decision-making power of local staffing groups (Eckhard 2019). Despite the theoretical principal-agent problems and practitioners' concerns, empirical evidence of a more corrupt behavior of national IO staff is scarce. In a rare attempt to empirically assess the question, Heinzel (2022a) found fewer "red flags" for potential corruption in World Bank procurement processes that were overseen by national staff compared to processes managed by international staff.

Another potential downside of relying on national staff might be the presence of biases towards specific ethnic, political, or religious groups in the host country that might affect local staff's decision-making. Findings from Kosovo illustrate such biases to have affected peacekeeping missions in the past (Eckhard 2014). Further, in a survey study national and international staff from 22 IOs found "impartiality to local groups" to be a main contribution of international staff

[1] For example, World Bank TTLs rotate every 3-5 years (Denizer, Kaufmann, and Kraay 2013).

[2] In the UN system, the Noblemaire principle of 1921 states that financial compensation of international bureaucrats should be at a competitive level compared to the world's best paid domestic civil service (Ogwezzy 2016).

to IO field missions (Eckhard and Parizek 2022). Empirical evidence for such behavior is yet scarce and limited to case studies (Platteau 2004; Takasaki 2011).

Despite these potential risks, local recruitment is also expected to provide substantial benefits, due to their deeper understanding of the local context, the culture, customs, and not the least the language of the host country. IO projects are complex endeavors that require high-quality access and local knowledge (von Billerbeck et al. 2024). Accordingly, a lack of understanding of the local context can have highly detrimental effects on outcomes (Campbell 2018; Easterly 2002). Even though organizations can also gain local knowledge when their international staff stays in a country for extended periods, the mentioned mobility requirements most IOs impose on their international staff limit this option (Heinzel and Liese 2021). National staff is meant to fill the knowledge gap of the IPS (Eckhard 2019; Kullenberg 2016). The NPO category in the UN system was specifically established for tasks that "required national knowledge and experience and therefore could not be carried out as effectively by international professionals" (ICSC 2009: 3). In further theorizing the knowledge sharing mechanisms of the local staff, Eckhard (2021: 296, 295) describes it as "the process by which bureaucrats—who possess advanced knowledge about a social environment which is affected by policy—share information and broker relations between their organization and that environment." In that vein, the NPOs are expected to support their organization by providing "informational knowledge (facts about culture, history, politics) and relational knowledge (how people interact)" about the local environment.

Case studies have collected numerous accounts from international staff highlighting how vital the *informational knowledge* expertise of local staffers is for the functioning of respective IO missions (Autesserre 2014; Coleman 2024; Eckhard 2021). As one UN IPS phrased it in a recent interview study: "Without local peacekeepers, there is no possibility to work in the country… Because you do not know the reality of the country, you don't know the country context. You don't know the dynamics. You don't know the local culture" (Coleman 2024: 23). This sentiment is widely prevalent among UN peacekeepers; when asked about their main focal points for collecting information on the local context, 66% referred to national colleagues (Eckhard 2021). Much of this local knowledge is not easily transmitted to headquarters or even verified, as it often falls into categories such as rumors and tacit knowledge related to local customs (Honig 2018, 2019).

Beyond country-specific expertise, local staffers also provide *relational knowledge* to enhance the access of IO staff to stakeholders and increase the trust of local partners (Autesserre 2014; Campbell 2018; Coleman 2024; Sundberg 2019). This access includes identifying the right focal points (Coleman 2024). In the context of Tanzanian aid projects, Sundberg (2019) notes that local staff often work in the relevant sector for extended periods and are well-connected to government agencies, other donor organizations, and civil society, thereby providing fruitful access to a broad range of stakeholders. Additionally, the persuasion of project opponents is easier when staff have country experience, as they can leverage their existing personal and professional relationships to foster support for new projects. Studies suggest that when recipients trust World Bank staff, they will be more open to dialogue and persuasion, and more inclined to adhere to their obligations in a project (Bazbauers 2019; Ika, Diallo, and Thuillier 2012; Lannon and Walsh 2020). By interpreting customary behaviors and translating across cultural barriers, local staffers can also mitigate prejudice towards IO staff and activities in dialogue situations with local communities (Campbell 2018; Eckhard 2021), and vice versa (Coleman 2024; Eckhard 2021).

In summary, local staffers engage in gathering context-specific information and facilitating local access and dialogue. Therefore, with substantial safeguards against local capture in place and no clear evidence of more corrupt behavior by national staff, the benefits of local embeddedness lead us to expect that:

H_1: The prevalence of national staff in IO country offices improves project performance.

National Staff in Fragile State Contexts

Some scholars have pointed out that the extent to which the unique assets provided by national staff translate into higher project success depends on multiple factors related to both the IO and the implementation country (Honig 2020; Marchesi and Masi 2021). As expertise and relational knowledge are likely the main contributions of local staff, their impact depends on how much of this knowledge is required and can be utilized. Honig (2018, 2019) discusses this point in detail, finding that local expertise is particularly relevant in countries with high levels of state fragility.

High fragility implies a low availability of standardized data and rapidly changing environmental conditions during implementation, making remote planning and blueprint approaches less suitable (Marchesi and Masi 2021). In line with this argument, multiple

scholars have discussed the potential downsides of blueprint project designs and have promoted more adaptive approaches (Booth 2015). Honig highlights the importance of understanding uncertainty and legibility: "Environments vary with regard to legibility—the extent to which they can be understood from a distance. More unpredictable environments are also likely to be less legible" (Honig 2019: 177).

That addressing this unpredictability will likely bring performance returns is highlighted by the finding, that WB projects that experience course corrections in form of a project restructuring during implementation, perform better than projects that are implemented as originally planned (Denizer et al. 2013). The WB apparently acknowledges the need for local expertise in uncertain contexts through their project design decisions. Marchesi and Masi (2021) find that WB projects are more likely to be implemented with subnational partners, who offer greater proximity to and knowledge about local beneficiaries and conditions when information availability and data quality in the host country are poor. While Honig emphasizes local knowledge (2018, 2019, 2020), this also applies to the increased need for local access in uncertain environments. While often understood as necessary to identify the right focal persons (Coleman 2024; Ward 2021), in highly fragile contexts, access can literally mean staff presence in potentially dangerous areas. In the field of humanitarian aid, there is a noted "rise of remoteness" (Tammi 2022: 968), where humanitarian IOs increasingly withdraw international staff from risky locations and shift operational responsibilities to local staff and partner organizations. In such contexts, local staff is therefore potentially a prerequisite for functioning on-site operations. Anecdotal evidence from countries like Afghanistan during the 2010s shows that international development agencies operated local offices exclusively with local staff, while providing international oversight from outside the country (Azimi et al. 2023).

Given their extensive contextual expertise, we expect that local staff are particularly beneficial in fragile environments where implementation conditions change frequently, and reliable standardized data is scarce. Therefore, we hypothesize that the relationship proposed in H_1 is moderated by the fragility of the country context.

H_2: In more fragile recipient countries, IO project performance benefits to a greater extent from national staff.

Empirical Analysis

We investigate our hypotheses using a mixed-methods design (Schoonenboom and Johnson 2017; Tashakkori and Teddlie 2003). First, we conduct a quantitative analysis drawing on a

combination of data on the staffing structure of 172 IO country offices and 5,176 evaluations of projects that were implemented at these duty stations, to test out hypotheses. Second, we present findings from a qualitative case study that validates the correlational results and explores the hypothesized causal mechanisms.

Quantitative Evidence

Our quantitative analysis seeks to empirically test the hypotheses that projects with greater involvement of local IO staff benefit from their country-specific expertise and enhanced access to local stakeholders (H_1), with the country context acting as a moderator in this relationship (H_2). This approach follows a y-centered research design (Ganghof 2005), focusing on whether the independent variable exerts an effect rather than explaining the full variance of the dependent variable. For the dependent variable, we make use of a text mining method recently developed by Eckhard et al. (2023) to measure the performance of 5,176 projects from five different UN system IOs based on their publicly available evaluation reports. For the independent variable, we draw on an original data set of UN system staff statistics that reports the staffing composition of all 1658 offices (at the subnational level) of 45 UN system IOs. We also control for a range of alternative explanations derived from the literature as discussed below. We apply multilevel and fixed effects regression models to enhance validity and generalizability of our findings and facilitate causal attribution.

Dependent Variable

The construction of our dependent IO project performance measure followed the method described in detail by Eckhard et al. (2023). The authors developed and validated a method to quantitatively evaluate the performance of IOs using the text of the evaluation reports. These reports assess the performance of various organizational activities but often do so in qualitative, text-based formats that are not straightforward to analyze quantitatively. To address this, the authors fine-tuned the BERT (Bidirectional Encoder Representations from Transformers; originally introduced by Devlin et al. 2018) large language model, a deep learning technique for natural language processing, to automatically classify sentences within these reports as positive, negative, or neutral assessments of the organization's performance. From these classifications, Eckhard et al. (2023) calculated the proportion of positive sentences per report, which served as a novel quantitative performance measure. The measure has been shown to be valid using content, convergent and construct validation (Adcock and Collier 2001; Eckhard et al. 2023).

We collected all evaluation reports available in English[3] as well as relevant meta data to the respective projects using web scraping of the evaluation reports databases for the following five UN system IOs: Food and Agriculture Organization of the United Nations (FAO), International Labour Organization (ILO), United Nations Development Programme (UNDP), United Nations Children's Fund (UNICEF) and World Food Programme (WFP). These IOs work predominantly in the sectors of humanitarian and food aid (UNICEF, WFP, FAO) as well as development aid (UNDP and ILO), which constitutes sectors in which evaluation practices are well developed and standardized. All included IOs are sizeable in budget and human resources and implement a broad range of projects in partner countries. For a summary of some of the characteristics of the included IOs see Appendix III. The case selection follows the diverse case logic (Seawright and Gerring 2008) which mitigates the risk of selection bias linked to a more homogenous sample. At the same time, all IOs are members of the UN Evaluation Group and subscribe to the same evaluation policy (UNEG 2016), which allows comparing the evaluation results.

In total 7.561 reports could be collected. We then implemented a textual preprocessing pipeline to extract the texts from the PDF files, remove unnecessary information like title pages, content tables and page numbers (wherever possible) and split the reports into sentences. We then used the BERT model of Eckhard et al. (2023) to classify the sentences of each evaluation report. In total all included reports contain roughly 9 Mio sentences. In line with Eckhard et al. (2023) we calculated the share of positive sentences of all non-neutral sentences as our project performance measure. The positivity score P_i for report *i* is calculated as the number of positive sentences divided by the number of non-neutral sentences:

$$P_i = \frac{S_{pos,i}}{S_{pos,i} + S_{neg,i}}$$

P_i takes a value between 0 and 1, where 1 indicates that all non-neutral sentences in the report are positive and constitutes the performance variable for this study. The performance value distribution follows exactly theoretical expectations, being normally distributed with a moderate positive bias as illustrated in Figure 2. A similarly sized positive bias is well documented for IOs and bilateral donors that publish quantitative ratings (Honig 2018).

[3] The current version of the BERT classifier does only work with English language reports. Due to resource limitations, a translation step was not included in the data processing pipeline.

The resulting reports dataset covers projects ended between 1995 and 2024. Unfortunately, the distribution of reports is very uneven between the IOs, with the majority coming from UNDP (4,592) and UNICEF (1,842), while ILO (650), FAO (248) and WFP (229) accounting for only small shares of the total included reports. We only collected reports in English language and disregarded those in French, Spanish and Portuguese due to a lack of resources to translate them.[4] The geographical distribution of the projects in the dataset is illustrated in Figure 1 which shows that especially Western African and South American countries are currently somewhat underrepresented in the dataset while South-East Asian Countries and the Middle East are overrepresented. This geographical bias is at least to some extent the consequence of the current exclusion of non-English reports.

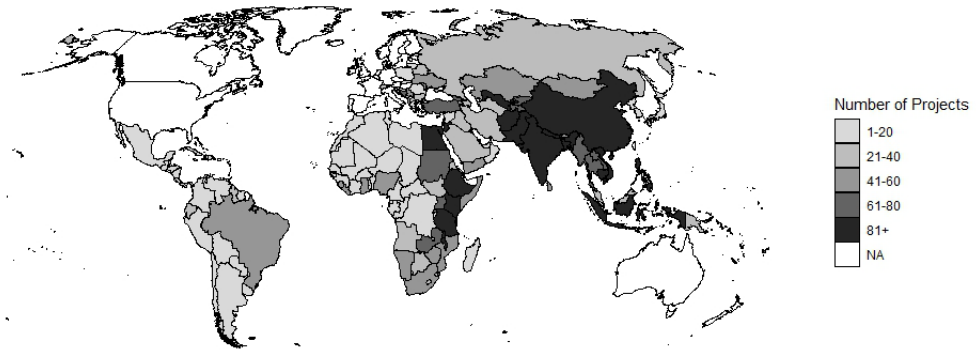


Figure 1: Geographical distribution of projects

Independent Variables

As mentioned above, staff members of the UN system fall in one out of different categories namely International Professional staff (IPS), National Professional Officers (NPO), General Service (GS) and Field Services (FS)[5]. Officials in the IPS usually make a lifelong career in the UN system outside their home country, often by serving in different IOs or at least at different duty stations within one IO (Eckhard and Steinebach 2021). NPO staff perform different technical and professional functions similar to those of IPS but are recruited locally and do not necessarily pursuit long-term careers within the UN system. UNICEF defines it like this: "Broadly speaking, the requirements for the N(ational) O(fficer) staff category are similar to

[4] We have 2,261 reports in French, Spanish or Portuguese (1,395 UNDP, 604 UNICEF, 180 ILO, 40 FAO, 42 WFP. The reports have approx. 200,000 characters each. The cost for translation with an up-to-date translator API such as DeepL API, would cost about four Euros per report and consequently over 9,000 Euros for all non-English reports. Furthermore, the translation per report takes about 30 minutes.

[5] As the category of FS mainly applies to specialized military, security, and logistics personnel, which are not hired by any IO from our analysis, this category will not be relevant in our analysis.

those of International Professional staff (IPS), with a more pronounced emphasis on in-depth knowledge of the local setting, institutions and partners" (UNICEF 2024). Employees of this group are specifically hired to contribute local expertise to IO field offices. NPO ranks are roughly equivalent to the ones of IPS and have similar job requirements (UNICEF 2024). GS describes a broad category of locally recruited "non-professional grades" (UNCEB 2023). GS staff performs administrative, building maintenance, and other support tasks and is hired locally based on fixed-term appointments (UNDP 2024; UNICEF 2024). Hence, while GS might bring in some local expertise, the predominant staff category to provide the required local knowledge are the NPOs. We will account for the effect of both local staff categories separately.

We retrieved personnel data for all UN system offices by staff category for the years 2007 till 2022 from the annually published UN Chief Executives Board for Coordination (UNCEB) "Personnel Statistics" reports (e.g. CEB/2023/HLCM/HR/4). The unit of the resulting dataset is on the level of duty station per IO and year. Regarding the five IOs that are subject to our analysis, the dataset contains a total of 18,755 entries for 998 unique office locations in 185 countries across five IOs and 16 years. The number of office locations by IO can be seen in Appendix III.

To attribute office locations to evaluation reports and thus to projects (our unit of analysis), we first compiled a list of office locations by country and year for each IO. Using regular expressions, we searched for these location names within the evaluation report texts. The frequency of location mentions in each report was used as a proxy for the involvement of the respective office. All locations mentioned in a report were included in further analysis[6]. To adjust for the potential overrepresentation of capital cities, which might be mentioned due to government-related activities rather than project implementation, we applied a devaluation factor of 0.7 to mentions of capital cities.

To account for the involvement of various duty stations mentioned in evaluation reports, we calculate relevance scores for each report location. The relevance score for each duty station d_k in report *i* is determined by the following formula:

$$\text{relevance}_{\text{i,d}_\text{k}} = \frac{\text{mentions}_{\text{i,d}_\text{k}}}{\text{total\_sentences}_\text{i}}$$

Where:

- $\text{mentions}_{\text{i,d}_\text{k}}$ is the number of times duty station d_k is mentioned in report *i*
- $\text{total\_sentences}_\text{i}$ is the total number of sentences in report *i*

[6] The maximum number of stations found in a report was 11.

Even though mere searching for location mentions is not an ideal approach we consider it the most feasible solution considering the limitations we are facing. As our text corpus consists of evaluation reports, the texts contain usually frequent mentions of office locations in the context of evaluator field visits. These field visits are likely a good proxy for the presence of relevant stakeholders as well as implementation sites. For example, the evaluation report on the UNDP "Sudan Mine Action Capacity Building and Development Project" (Paterson and Bohle 2008) contains 14.7 weighted mentions of Khartoum and 15 of Juba in the report comprising 1012 sentences. The first mention of both cities is in the acknowledgment section, where the responsible officers at both locations are thanked for their support during the evaluation. At times the office structure is particularly mentioned[7] or is mentioned in the context of the travel itinerary included in the evaluation. Khartoum also gets mentioned in the context chapter which highlights the necessity to devalue capital city locations[8].

Using the calculated relevance scores staff numbers by category were integrated into the dataset based on the IO, year, and duty station. Using the relevance scores, we calculate the weighted local staff share for each report *i*:

$$\text{LS\_share}_{\text{i}} = \sum_{\text{k}=1}^{\text{d}_{\text{i}}} \left(\frac{\text{N}_{\text{NPO,i,d}_{\text{k}}} + \text{N}_{\text{GS,i,d}_{\text{k}}} \times \text{relevance}_{\text{i,d}_{\text{k}}}}{\text{N}_{\text{TOTAL,i,d}_{\text{k}}} \times \text{relevance}_{\text{i,d}_{\text{k}}}} \right)$$

Where:

- $\text{LS\_share}_{\text{i}}$ is the weighted share of all national staff (NPO and GS) for report *i*.
- d_{i} is the number of duty stations involved in report *i*.
- $\text{N}_{\text{NPO,i,d}_{\text{k}}}$ and $\text{N}_{\text{GS,i,d}_{\text{k}}}$ are the number of NPO and GS staff respectively at duty station d_k for report *i*
- $\text{relevance}_{\text{i,d}_{\text{k}}}$ is the relevance score of duty station d_k for report *i*.
- $\text{N}_{\text{TOTAL,i,d}_{\text{k}}}$ is the total number of staff at duty station d_k for report *i*.

The weighted staff shares for NPOs and GS were also computed separately to investigate the effects of the individual local staff categories on performance. By multiplying the number of staff at each duty station by its relevance score, and then normalizing by the total staff (also weighted by relevance), we obtain an approximative measure of the proportionate involvement of the local staff for the report. The resulting staff share measure has a value from 0-1 with 1 indicating a greater share of the respective staff category involved in the project. The

[7] "UNDP is the lead agency within UNMAO for supporting the development of mine action capacities. The UNDP project is managed by a Senior Technical Advisor (STA) based in Khartoum, with a Technical Advisor (TA) in Juba and two national programme officers." (Paterson and Bohle 2008: 10).

[8] "While the principal conflict has been between the national government in Khartoum and Sudan People's Liberation Movement (SPLM7) in the South, insurgencies have also occurred in the East and (continuing today) in Darfur." (Paterson and Bohle 2008: 3).

distributions of local staff (combined measure) can be seen in Figure 2, the distributions by individual staff category are included in Appendix IV.

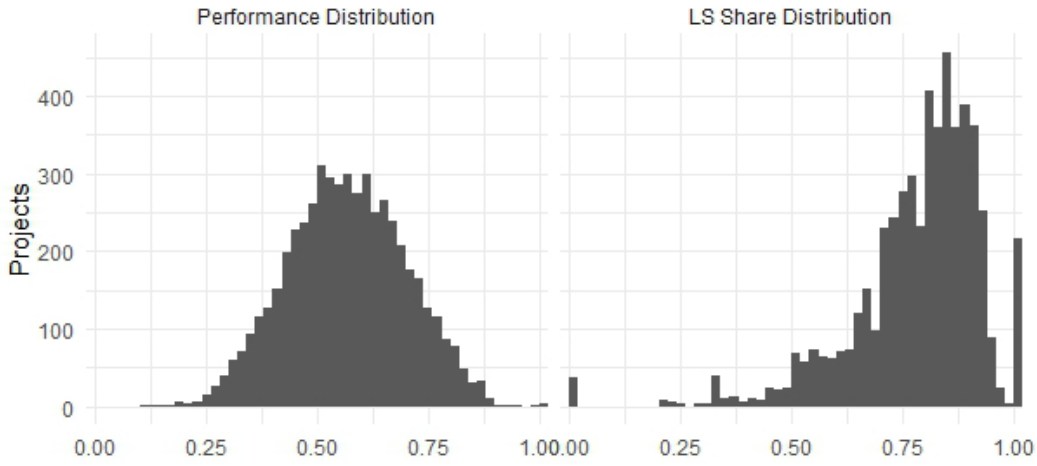


Figure 2: Distributions of Performance and Local Staff

Unfortunately, the temporal scope of our staff data is more limited than that of our evaluation dataset, as complete staff reports including all staff categories are only available since 2007. Our analysis is therefore limited to 5260 reports. After cleaning for other missing data (e.g. missing country and time data) and extreme outliers[9] 5176 reports remain in the dataset.

Moderator and Controls

To assess the level of uncertainty in the implementation countries, we follow Honig (2019) who uses the level of state fragility as a closely related concept. The Fund for Peace (2017) developed the Fragile States Index (FSI) to assess social, economic, and political pressures on states using 12 composite indicators derived from over 100 sub-indicators like demographic pressures, refugees and internally displaced persons, group grievance, human flight, uneven economic development, poverty and economic decline, state legitimacy, human rights and rule of law, security apparatus, and external intervention. The Fragile States Index score ranges from 0–120, is normally distributed, and can be interpreted as follows: sustainable (0–30), stable (30–60), warning (60–90), and alert (90–120) (Fund for Peace 2017).

Besides, we control for some confounders on the project and country level. Compared to studies that rely on WB data, the project level meta-data available to this study is unfortunately limited. At the project level, we include the length of the evaluation report, measured in sentences, as

[9] We remove outliers in terms of our main local staff variable as well as our dependent staff variable which are driven by reports with a very low number of non-neutral sentences and IO locations with extremely small staff numbers (one or two).

an approximate indicator of project complexity. We also incorporate a binary measure for joint evaluations as additional stakeholders might increase the risk of positive bias (Malik and Stone 2018) and co-funding organizations have been shown to reduce effectiveness (Winters 2019), which we obtain from scraped websites. Additionally, we aimed to generate supplementary text-based variables. However, utilizing the entire reports for this purpose would be highly labor-intensive and imprecise, especially given the input token limitations of standard language models (e.g., BERT, with a 512-token limit). To address this, we employed the OpenAI GPT-3.5 turbo API to produce detailed summaries of approximately 400 words. We submitted entire reports, or up to 10,000 words for longer documents, to the API.[10] These summaries, containing aggregated information from the reports, were then used in conjunction with the titles to create the additional variables (using dictionary approaches) such as a binary measure indicating the timing of the evaluation (mid-term or ex-post). Moreover, we assign sector codes based on the Creditor Reporting System (CRS) of the OECD, utilizing a similar dictionary approach. Furthermore, we include a measure of overall office decentralization in the respective country measured as number of offices an IO holds in the respective year and country. Researchers commonly include measures on governance capacity and institutional quality (Ferry, Hafner-Burton, and Schneider 2020; Heinzel 2022a; Honig 2020). Hence, at the country level, we include the political regime using the V-Dem liberal democracy score and the total official development assistance (ODA) received by the host country (log-transformed). Also, GDP per capita, and population size (log-transformed), are included to control for the scale of the country's economy and population.

Findings

We report the main findings on the hypothesized relationships below. A summary of central variables of the dataset is reported in Table 1. Several descriptive visualizations, including the correlation of local staff share with performance by IO, and the distributions of the staff shares are available in Appendix III-V.

Table 1: Dataset Summary

Variable	*Mean*	*SD*	*N*
Performance (0-1)	0.567	0.130	5176
LS share (0-1)	0.787	0.134	5176

[10] Tokens were truncated at 10,000 words to adhere to the API's 16,000-token limit per request and response. This buffer was necessary due to the Byte Pair Encoding (BPE) method used for tokenization, which results in a higher token count than the actual word count. Although longer reports were not fully processed, the inclusion of executive summaries ensured the model accessed the most critical information.

NPO share (0-1)	0.301	0.138	5176
GS share (0-1)	0.486	0.133	5176
Report Length (sentences)	1222.321	716.990	5176
Joint Evaluation (binary)	0.095	0.294	5176
Mid-Term Evaluation (binary)	0.151	0.358	5176
Number of Offices	2.29	2.54	5176
Liberal Democracy (0-1)	0.307	0.196	5076
State Fragility (0-120)	81.198	15.013	5138
GDP per Capita (current USD)	4413.34	7063.78	5092
Population (log)	16.697	1.864	5176
ODA Total (log)	20.304	1.314	4741

National Staff and Project Performance

In line with the y-centered design, our goal is not to explain performance comprehensively but to assess whether the share of local staff has a meaningful effect on it. We report both a fixed effects (FE) and a multi-level model. Fixed effects models are preferred when the primary goal is to control for unobserved, time-invariant characteristics within each group (such as countries or IOs), providing consistent and unbiased estimates for within-group variations. We therefore use OLS regressions with recipient country and IO fixed effects in our analysis. IO fixed effects are necessary to absorb unobserved factors related to the IO evaluation processes and standards. However, fixed effects models absorb the variation of group-level variables that do not vary within the groups. As some smaller countries in the dataset have only one country office, their within-country variance in staff composition varies only over time. Yet, time-variant unobserved characteristics might contribute to changes in performance. Therefore, we also report findings from a multilevel model. In addition, the results from the fixed effects regression may be biased due to the violation of the observation independence assumption. Thus, we consider the way in which IO projects are nested into clusters, by including random effects for each cluster-variable into our multilevel models. Both our fixed and multilevel models include only the main predictor and cluster variables in the first model specifications (models 1 and 4).

In a second step we add all project-level covariables (models 2 and 5) and in a third additional country-level covariables (models 3 and 6). As the latter do not vary on the country_year level, their inclusion only accounts for temporal changes of these variables within countries in the fixed effects models. To account for sectoral differences, we include sector dummies (sectors are not exclusive within projects). In the fixed effects models, we include four-year interval fixed effects to account for some of the unobserved time-variant characteristics. In line with similar studies, the standard errors in the fixed effects models are clustered by evaluation year to correct for correlated errors arising from evaluations conducted within the same year

(Denizer et al., 2013; Heinzel, 2022; Heinzel and Liese, 2021). The results of all regressions are reported in Table 2.

Our local staff measure shows a consistently positive and highly significant positive effect across model specifications. Mid-term evaluations indicate a significantly less positive project performance than their ex-post counterparts. Longer reports might be significantly more negatively rated, but the results are unclear when time-invariant country characteristics are more rigorously controlled by including fixed effects.

Increasing state fragility leads to lower performance and this finding is consistently significant across models and meaningful in size. Projects implemented in more populous and richer countries (in terms of GDP per capita) are also significantly more likely to perform well. Our findings with regards to the included controls are consistent with previous findings from project level studies, as well as our explained variance[11] which falls in the range of $R^2 = 0.143 – 0.176$ (Denizer et al. 2013; Honig 2019, 2020; Heinzel and Liese 2021).

Table 2: Fixed Effects and Multi-Level Model Results

	Dependent variable: Performance					
Variables	*Fixed Effects Models*			*Multi-Level Models*		
	(1)	(2)	(3)	(4)	(5)	(6)
Local Staff	0.092***	0.088***	0.074***	0.131***	0.130***	0.147***
	(0.022)	(0.023)	(0.020)	(0.015)	(0.015)	(0.017)
Mid Term		-0.346***	-0.333***		-0.321***	-0.325***
		(0.044)	(0.043)		(0.038)	(0.040)
Joint Evaluation		-0.056	-0.054		-0.083	-0.040
		(0.041)	(0.042)		(0.048)	(0.049)
Report Length		-0.036**	-0.034*		-0.036*	-0.039*
		(0.013)	(0.013)		(0.014)	(0.015)
Office Decentralization		-0.031	-0.034		-0.030	-0.061***
		(0.021)	(0.022)		(0.016)	(0.019)
Fragility Index			-0.137			-0.112***
			(0.093)			(0.028)
Democracy			-0.073			-0.046*
			(0.063)			(0.020)
ODA Total			0.042			0.128***
			(0.035)			(0.024)
GDP per Capita			0.207**			0.096*
			(0.078)			(0.042)
Population (log)			1.559**			0.084**
			(0.595)			(0.023)

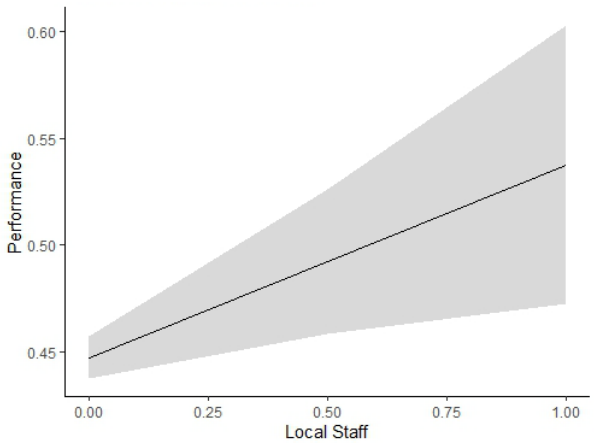
[11] Log-Likelihood and Akaike Information Criterion (AIC) indicate better model fit of the fixed effects model, Bayesian Information Criterion (BIC) conversely prefers the multilevel models.

Constant				-0.013	-0.050	-0.075
				(0.034)	(0.045)	(0.044)
Country and IO FE	Y	Y	Y	N	N	N
4-Year-Interval FE	Y	Y	Y	N	N	N
Country, IO, Year RE	N	N	N	Y	Y	Y
Sector Dummies	N	Y	Y	N	Y	Y
AIC	14353.4	14227.2	12440.4	14390.2	14276.4	12539.4
BIC	15342.8	15386.9	13475.1	14429.6	14486.0	12777.1
Log-Likelihood	-7025.7	-6936.6	-6059.2	-7189.1	-7106.2	-6232.7
Observations	5,176	5,176	4,566	5,176	5,176	4,566
R^2	0.143	0.174	0.176			

Standard errors in parentheses; errors of FE models clustered at evaluation year; $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$

To improve comparability, all included variables are z-standardized. To better interpret the size of our main effect, we calculated the marginal effects for Model (2). We use this model as it includes all project level variables but at the same time does not include country level variables, that are largely controlled by the inclusion of country fixed effects. By using this model, we mitigate the risk of overfitting. In terms of increase every raise of local staff share by 25 percentage points expects an increase of positive assessments in a project report of 2.5 percentage points. As Eckhard et al. (2023) illustrated when comparing the text-based measure with human coded outcome ratings of the World Bank, the covered scope between 46% and 54% positive assessments in the text translates to a change from a "moderately unsatisfactory" to a "moderately satisfactory" outcome rating[12]. Hence, the effect of local staff share at a given project can be considered to have small but nonetheless meaningful impact on the assessment of project performance. We therefore find our analysis to support H_1.

Figure 3: Estimates of Marginal Effects



[12] The World Bank's Independent Evaluation Group uses a 6-point Likert scale to measure outcome performance.

National Staff in Fragile State Contexts

With the second hypothesis we expect IO projects in fragile countries to especially benefit from higher rates of local staff during implementation. To test this hypothesis, we include an interaction term between state fragility and local staff in our analysis. Table 3 examines the interaction between the share of local staff and state fragility on project performance. We exclude the 4-year-interval fixed effects in these models to avoid absorbing the temporal variance of the fragility measure (Honig 2019). Consistent with the findings reported in Table 2, projects with higher shares of local staff perform significantly better. Also aligning with the findings of Honig (2018), as fragility increases, performance significantly decreases. Contrary to our expectations, the interaction term between local staff and environmental fragility is neither substantial in size nor statistically significant. This suggests that while local staff generally improves performance, the benefits do not differ significantly between more fragile states and less fragile ones. This finding is still relevant as it implies that the advantages of employing local staff are consistent across various levels of state stability, reaffirming the overall hypothesis that local knowledge positively impacts project outcomes irrespective of the fragility context.

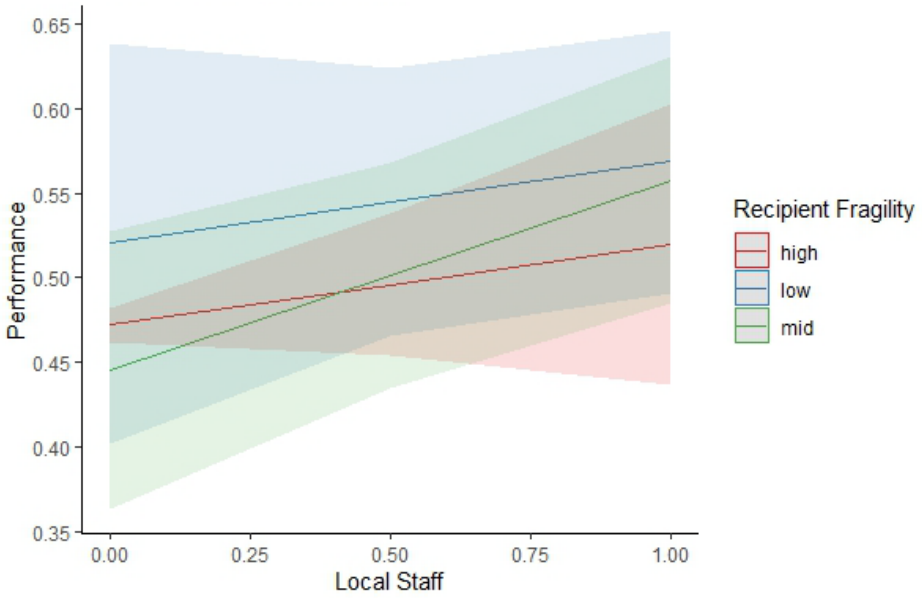
Table 3: Local Staff and Fragility Interaction

	Dependent variable: Performance		
Variables	(7)	(8)	(9)
Local Staff	0.073**	0.072**	0.072**
	(0.023)	(0.023)	(0.023)
Fragility Index	-0.367***	-0.390***	-0.229***
	(0.083)	(0.089)	(0.069)
Local Staff*Fragility	0.002	0.002	0.003
	(0.024)	(0.024)	(0.026)
Country and IO FE	Y	Y	Y
Sector Dummies	Y	Y	Y
Project Controls	N	Y	Y
Country Controls	N	N	Y
Observations	5,138	5,138	4,566
R^2	0.145	0.157	0.176

*S*tandard errors in parentheses, clustered at evaluation year; *p > 0.05, **p<0.01, *** p<0.001

To further investigate the potential of a non-linear relationship, we plot the marginal effects of local staff on performance by fragility level. The categorical variable for fragility level follows the measure of the Fund for Peace, which defines countries with a score from 0-30 as "sustainable", 30-60 as "stable", 60-90 as "warning" states, and 90-120 as "alert" states. As sustainable and stable countries are underrepresented in our dataset, we combine these two categories, as all countries falling into this group can be considered stable. We plot the marginal effects of the model (2) and plot the marginal effects by fragility level in Figure 3. Despite our theoretical expectation, the marginal effects appear to imply a non-linear relationship, as the effect of local staff is strongest in mid-fragility environments and weaker in both low fragility (i.e. stable) and high fragility environments. The finding, that staff has less of a positive effect in stable environments follows our expectations, as in highly stable environments, the benefit of local knowledge may be less pronounced because such environments already have robust data systems and institutional structures and therefore less need for soft knowledge and access. The latter is surprising and will be discussed more extensively below. Despite the interesting finding of potential non-linear relationship of the local staff and fragility interaction, none of our findings support our linear expectation stated in H_2.

Figure 4: Estimates of Marginal Effects by Fragility Level



Robustness

To assess the robustness of our findings we ran additional tests with the disaggregated local staff variables based on their categories. To address potential time-variant unobserved effects, we included 4-year time fixed effects. Standard errors of the fixed effects models are clustered by evaluation year. The results of the models are reported in Table 4. They are generally

consistent with the main findings presented in Table 2. Both individual staff categories positively correlate with higher performance, when controlling for sector and potential confounders. Overall, the effect National Professional staff has on performance is more consistently significant across model specifications and more substantial in size than the one of General Services staff. This supports our hypothesized causal mechanism, as NPOs are specifically hired for the local knowledge it provides. It likely also highlights the importance of staff discretion, as NPO are roughly equivalent to their international counterparts and have substantially greater decision-making competencies than their GS colleagues (UN WOMEN n.d.).

Table 4: Fixed Effects and Mixed Effects Regressions

	Dependent Variable: Performance							
Variables	*Fixed Effects Linear Model*				*Linear Mixed-Effects Models*			
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
NPO Share	0.062***		0.105***	0.086***	0.112***		0.159***	0.170***
	(0.017)		(0.022)	(0.021)	(0.015)		(0.018)	(0.021)
GS Share		0.011	0.075**	0.068*		0.019	0.103***	0.138***
		(0.018)	(0.028)	(0.027)		(0.016)	(0.018)	(0.020)
Constant					-0.013	-0.007	-0.052	0.061
					(0.029)	(0.033)	(0.044)	(0.032)
Recipient, IO, 4-Year FE	Y	Y	Y	Y	N	N	N	N
Recipient, IO, Year RE	N	N	N	N	Y	Y	Y	Y
Sector Dummies	N	N	Y	Y	N	N	Y	Y
Project Controls	N	N	Y	Y	N	N	Y	Y
Country Controls	N	N	N	Y	N	N	N	Y
Log-Likelihood	-6949.5	-6955.3	-6848.7	-6048.8	-7200.2	-7225.1	-7102.3	-6273.3
Observations	5,176	5,176	5,176	4,566	5,176	5,176	5,176	4,566
R^2	0.141	0.139	0.174	0.180				

Standard errors in parentheses; errors of FE models clustered at evaluation year; *p<0.05, **p<0.01, ***p<0.001

Qualitative Illustration

In the following, we present findings from an illustrative case study conducted in Lebanon in 2023. Although we cannot fully capture the depth of the case study data in this context, we believe these data effectively illustrate the mechanisms we have theorized. We aimed to select a 'typical' country case, meaning it was not expected to be extreme on either of the independent variables—local staff involvement and fragility[13] (Seawright and Gerring 2008). We chose Lebanon due to its longstanding presence of international organizations (IOs) and its political significance at the time of research. Between 2007 and 2022, 15 UN IOs maintained offices in Lebanon, including all the IOs examined in our quantitative study. We present findings from seven semi-structured, key-informant interviews (extracted from a broader dataset of 27 interviews) with international and local IO staff in Lebanon, specifically from the UNDP and the United Nations Interim Force in Lebanon (UNIFIL). These IOs were selected due to their widespread presence and broad range of projects that bridge the security-development nexus, thereby linking the literature on local staff (predominantly from peacebuilding) with our quantitative study (focusing on development and humanitarian IOs). Evidence presented here emerges from a three-week fieldwork period in Lebanon in September 2023. Interviewees were recruited using snowball sampling methods, leveraging the personal and professional networks of the broader research team for initial entry points. Access to UNIFIL was attained through a formal request, granted by the Deputy Head of Mission. Interviews were audio/video recorded with the consent of participants and transcribed verbatim using automatic transcription functions in Microsoft Word, Teams, and Zoom, followed by cleaning and verification by research assistants. Explorative data analysis[14] was based on the following research question: how do IO interviewees understand and evaluate the role of national staff in the performance of international organizations?

[13] The average Fragility levels between 2006 and 2023 is with FSI = 88,5 slightly higher than the average observed in the quantitative dataset but lies within our most common category “warning”. Lebanon’s security situation changed multiple times during the observed times-scope wherefore we expect some variance in fragility within this one country case. With respect to other included country level controls it reflects the countries in the quantitative dataset well, despite being smaller in population and richer in terms of GDP per capita (*see* Appendix II).

[14] Interviews have been analyzed using NVivo 14 software according to a descriptive, open-coding protocol (Saldana 2015) for the purposes of thematic analysis (Fereday and Muir-Cochrane 2006). Initial coding was data-driven and consisted of “meaning condensation” (Kvale 1996: 192), codes were then categorized and organized into parent and child hierarchies to identify connections and relationships in the data (Baralt 2011).

This study presents evidence that national staff positively impact IO effectiveness based on the two key impact mechanisms described earlier: provision of contextual knowledge (see von Billerbeck et al. 2024; Coleman 2024; Kullenberg, 2016), and relations with host communities (see Eckhard 2021; Campbell 2018; Autesserre 2014).

As hypothesized, local staff positively impact IO performance by enhancing the situational awareness and contextual knowledge of their teams through their breadth of knowledge, depth of understanding, and awareness of the broader local climate. Due to their rootedness and often long careers with IOs (Sundberg 2019; Coleman 2024), local staff possess a superior breadth of knowledge compared to their international colleagues, including a deeper understanding of history, memory, and perspective.

> We have also people who are not rotating in and out. [...] my colleagues' analysis was very, very rich, because they said, you know, five years ago, it was like that, now it's like that, and this guy is like that because he comes from that (UNDP 4).

> I do value the importance of national staff, because I've been here long, but they've been even longer. So sometimes I do say things, or I try to implement certain projects, and they say, 'maybe it's not working out' and 'maybe we should be…'. (UNIFIL 14).

These quotations highlight the value of historical perspective and institutional memory, alongside the capacity of local staff to understand the complexities and sensitivities of their own communities (depth of knowledge). Together, these attributes make their voices and perspectives essential for mitigating tensions or avoiding detrimental outcomes (Campbell 2018; Easterly 2002), as well as for effective conflict-sensitive analysis and implementation.

> [W]hatever you receive here in terms of information, I realize that it can be analyzed in 20 different ways […] when you confront with everyone in the office [office is 80% national staff], everyone has a different understanding (UNIFIL 14).

Diversity among local staff produces a microcosm of host societies, enabling IOs and international staff to develop more inclusive and representative understandings to inform their decision-making and activities. In a complex and politically fragmented society, rather than generating confusion, diverse perspectives positively contribute to the effectiveness and appropriateness of IO projects. A common sentiment among participants was that local staff help IOs to "get the pulse of the population" (UNIFIL 15) and comprehend "the situation in the Lebanese mindset" (UNIFIL 17). This sense of the local climate can function as an informal early warning system (Cassin 2022), contributing to IO adaptivity, preparedness, and responsiveness.

Local staff also positively impact IO performance by *facilitating relations* with host communities. This occurs through three main pathways: trust building, face-to-face interactions, and effective communications. Local staff are embedded members of their communities, with shared experiences and histories, meaning their capacity to understand, empathize and relate can help build trust between IOs and their partners (Sundberg 2019).

> Yeah, the people would feel more safe in having something, somebody from their community who is familiar, who knows the community, in order to discuss all these critical issues with them (UNDP 7).

These same attributes facilitate greater and more continuous face-to-face interaction with community members (Eckhard 2021), particularly for local field officers involved in implementation (Tammi 2022).

> [T]hey have this, continuous, you know, relationship building process with the people – meetings every day, coffees, you know, getting to know people, gathering information, discussing, following up on the news, being up to date with what's happening. This is a continuous work. It's happening online, offline, during working hours, outside working hours (UNDP 5).

Together, these quotations reveal the capacity of local staff to access greater honesty, trust and openness through their daily, professional and personal interactions. Many IOs rely on partnerships with municipalities, civil society organizations, and local contractors for the implementation of their projects and mandates. Further, positive relations and acceptance is essential to the ability of IOs to work safely and effectively. The maintenance of positive relations is often connected to effective and appropriate communication. As an UNDP national staff member noted, "the most difficult part […] is] coaching the team on a day-to-day basis on reactions, on how to talk" (UNDP 5). This de facto training role has been widely observed in the literature (Baker 2019; Cassin 2022; Da Costa and Karlsrud 2013) and is an essential contribution to public perception, tension reduction/mitigation, cultural sensitivity, and the diplomatic capacities of IOs.

The interview evidence clearly illustrates the expected relationship between local staff and improved performance, thereby validating our quantitative findings in terms of causal attribution.

Discussion and Limitations

Even though our methods aim to address major shortcomings of previous studies by employing a mixed-method design, project-level data, and robust regression models, they are not without

limitations. One limitation concerns our measure of local staff involvement. Unlike previous studies, we sought to align office-level staff data with project-level outcome data. This attribution process, based on the text of the respective project evaluations, assumes that office locations are mentioned in the text and that the frequency of these mentions reflects the relevance of the location as an implementation site. As evaluation reports include besides information on project implementation also information on the evaluation itself such as itineraries of evaluators, our approach is likely more feasible in this context than alternative measures such a geographic proximity. Since our only available metadata regarding project location is at the country level, a fine-grained attribution based on geographical proximity (*see* Dreher, Lang, and Reinsberg 2024) would anyways not be possible. Moreover, the relationship between geographical proximity and actual office involvement in a specific project remains unclear, particularly in the absence of additional information about the nature of the office.

Another limitation lies with the reliance on our text-based performance measure. While the measure has been proofed valid when compared to theoretical expectations and human ratings (Eckhard et al. 2023), evaluations are still subjective and potentially influenced by external pressures (Azzam 2010; Jankauskas and Eckhard 2023; Malik and Stone 2018; Raimondo 2018). In this it does not differ from quantitative evaluation ratings, which are commonly used in comparable project-level studies (Assefa, Rivera, and Vencatachellum 2014; Bulman, Kolkma, and Kraay 2017; Denizer, Kaufmann, and Kraay 2013; Raimondo 2016; Shin, Kim, and Sohn 2017). All additional included datasets fall in the time after the commonly applied OECD-standards of the DAC criteria were published. The consistency of the evaluations as metric for performance is improved by the fact that all investigated IOs are part of the UN Evaluation Group and controlled using IO fixed (and random) effects.

While our findings show a consistent, significant and meaningful positive effect of local staff on project-level performance (H_1), they do not support the common expectation that increased local staff presence enhances performance more in fragile country environments (H_2). This discrepancy may be attributed to several factors:

First, as previously discussed, the discretion of field agents is often considered crucial for the effective application of local expertise. Campbell (2018) has noted that field agents may sometimes need to act against directives from headquarters to achieve better performance outcomes. This necessitates a certain degree of de facto autonomy. Honig (2021) explored this in the context of World Bank projects, finding that field agent autonomy, rather than mere presence, is the primary driver of improved performance under environmental uncertainty. The

investigated IOs might not grant sufficient autonomy to IO field offices for us to observe stronger performance improvements through local staff in fragile contexts.

Second, our theoretical expectations on how local knowledge can better contribute to fragile contexts predominantly rely on findings from peacekeeping (Autesserre 2014; Campbell 2018; Coleman 2024; Eckhard 2019), World Bank high-level employees (Heinzel 2022b), or World Bank partner organizations (Marchesi and Masi 2021). These cases may not be representative of our conceptualization of local staff in economic and humanitarian aid IOs from the UN system. Peacekeeping missions are, by definition, implemented only in highly fragile and uncertain environments, thus findings from this body of research might offer limited insights into contexts of lower fragility. Regarding the findings from the World Bank (Honig 2020; Heinzel 2022; Marchesi and Masi 2021), they focus on the qualifications and autonomy of high-level staff who possess substantial discretion, particularly following reforms in implementation strategy that grant greater autonomy to country directors (Honig 2020). These findings cannot be expected to be directly translated to lower and mid-rank project employees that are mostly included in our local staff measure. A possible explanation of the curvilinear relationship we observe between country fragility and the effect of local staff, is therefore, that local staff autonomy is more restricted in more fragile environments.

Conclusion

This study provides a mixed-method examination of the role of locally recruited staff in enhancing the project performance of international organizations, with a particular focus on UN system operations. Our research integrates quantitative data from 5,176 evaluation reports from five UN IOs with office-level staff data and a qualitative case study from UNIFIL and UNDP operations in Lebanon.

Our quantitative analysis indicates that a higher proportion of local staff correlates positively and significantly with better project performance. This finding holds across various modeling approaches and remains significant when accounting for report- and country-level factors. The positive effect is especially pronounced for National Professional Officers, who are hired for their country-specific expertise and hold greater decision-making power. Contrary to expectations, the positive impact of local staff does not increase linearly with greater country fragility; instead, it decreases in the most fragile countries. This could be due to our choice of measurement, the diverse effects of fragility on staffing, or limited discretion for staff in more fragile contexts.

The qualitative findings from Lebanon show that the positive effect of local staff on IO effectiveness can be explained through two main mechanisms: contextual knowledge and relations with host communities. Their deep cultural and historical understanding enables them to navigate complex socio-political environments effectively, facilitating smoother project implementation and fostering stronger relationships with host communities.

This research supports a strategic shift towards greater reliance on local expertise, demonstrating its benefits for IO performance. The fact that UN IOs increasingly rely on international staff—who constitute the fastest-growing personnel category in the UN system—might negatively impact the overall project-level performance of these IOs. Furthermore, this study highlights the potential of text-based methods, such as the LLM method used here, to advance research in international public administration.

References

Adcock, Robert, and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95(3): 529–46. doi:10.1017/S0003055401003100.

Ashton, Louise, Jed Friedman, Diana Goldemberg, Mustafa Zakir Hussain, Thomas Kenyon, Akib Khan, and Mo Zhou. 2023. "A Puzzle with Missing Pieces: Explaining the Effectiveness of World Bank Development Projects." *The World Bank Research Observer* 38(1): 115–46. doi:10.1093/wbro/lkac005.

Assefa, Yemesrach, Oscar P. Rivera, and Désiré Vencatachellum. 2014. "Macro and Micro Determinants of Project Performance." *African Evaluation Journal* 2(1): 9 pages. doi:10.4102/aej.v2i1.86.

Autesserre, Séverine. 2014. *Peaceland: Conflict Resolution and the Everyday Politics of International Intervention*. Cambridge: Cambridge University Press.

Azimi, Abbas, Fahim Abed, Mary Bulman, Bamdad Esmaili, Ben Heubl, Lena Kampf, and Martin Kaul. 2023. "Das Verspechen." *Süddeutsche Zeitung*. https://www.sueddeutsche.de/projekte/artikel/politik/afghanistan-krieg-ortskraefte-bundesregierung-hilfe-e752694/ (August 5, 2024).

Azzam, Tarek. 2010. "Evaluator Responsiveness to Stakeholders." *American Journal of Evaluation* 31(1): 45–65. doi:10.1177/1098214009354917.

Baker, Catherine. 2019. "Language Intermediaries and Local Agency: Peacebuilding, Translation/Interpreting and Political Disempowerment in 'Mature' Post-Dayton Bosnia–Herzegovina." *Journal of War & Culture Studies* 12(3): 236–50. doi:10.1080/17526272.2019.1644413.

Baralt, Melissa. 2011. "Coding Qualitative Data." *Research methods in second language acquisition: A practical guide*: 222–44.

Bazbauers, Adrian Robert. 2019. “World Bank Technical Assistance: Locating and Cultivating Sympathetic Interlocutors.” *Development Policy Review* 37(5): 638–56. doi:10.1111/dpr.12377.

von Billerbeck, Sarah, Katia Coleman, Steffen Eckhard, and Benjamin Zyla. 2024. “Local Knowledges in International Peacebuilding: Acquisition, Filtering, and Bias.” *International Studies Review*. https://centaur.reading.ac.uk/117256/ (July 22, 2024).

Booth, David. 2015. “Still Watering White Elephants? The Blueprint versus Process Debate Thirty Years On.” *Perspectives on politics, production and public administration in Africa, Essays in honor of Ole Therkildsen*: 11–25.

Bulman, David, Walter Kolkma, and Aart Kraay. 2017. “Good Countries or Good Projects? Comparing Macro and Micro Correlates of World Bank and Asian Development Bank Project Performance.” *The Review of International Organizations* 12(3): 335–63. doi:10.1007/s11558-016-9256-x.

Buntaine, Mark T., and Bradley C. Parks. 2013. “When Do Environmentally Focused Assistance Projects Achieve Their Objectives? Evidence from World Bank Post-Project Evaluations.” *Global Environmental Politics* 13(2): 65–88. doi:10.1162/GLEP_a_00167.

Campbell, Susanna P. 2018. *Global Governance and Local Peace: Accountability and Performance in International Peacebuilding*. 1st ed. Cambridge University Press. doi:10.1017/9781108290630.

Cassin, Katelyn. 2022. “‘The Lebanese Way’: Hybridization and Cultural Peacebuilding through’interfaces and Interchanges’ across the Peacekeeper-Local Divide.” phd. Université d’Ottawa/University of Ottawa.

Coleman, Katharina. 2024. “National Staff and the Protection of Civilians in UN Peacekeeping: Facilitative and Direct Agency.” *International Peacekeeping*.

Collier, Paul, and David Dollar. 2004. “Development Effectiveness: What Have We Learnt?” *The Economic Journal* 114(496): F244–71.

Da Costa, Diana Felix, and John Karlsrud. 2013. “‘Bending the Rules’: The Space between HQ Policy and Local Action in UN Civilian Peacekeeping.” *Journal of International Peacekeeping* 17(3–4): 293–312.

Dávid-Barrett, Elizabeth, and Mihály Fazekas. 2020. “Anti-Corruption in Aid-Funded Procurement: Is Corruption Reduced or Merely Displaced?” *World Development* 132: 105000. doi:10.1016/j.worlddev.2020.105000.

Denizer, Cevdet, Daniel Kaufmann, and Aart Kraay. 2013. “Good Countries or Good Projects? Macro and Micro Correlates of World Bank Project Performance.” *Journal of Development Economics* 105: 288–302. doi:10.1016/j.jdeveco.2013.06.003.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. “Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding.” *arXiv preprint arXiv:1810.04805*.

Dollar, David, and Lant Pritchett. 1998. “Assessing Aid-What Works, What Doesn’t, and Why.”

Dreher, Axel, Valentin Lang, and Bernhard Reinsberg. 2024. “Aid Effectiveness and Donor Motives.” *World Development* 176: 106501. doi:10.1016/j.worlddev.2023.106501.

Easterly, William. 2002. “The Cartel of Good Intentions: The Problem of Bureaucracy in Foreign Aid.” *The Journal of Policy Reform* 5(4): 223–50.

Eckhard, Steffen. 2014. *Bureaucratic Representation and Ethnic Bureaucratic Drift : A Case Study of United Nations Minority Policy Implementation in Kosovo*. Konstanz: Bibliothek der Universität Konstanz.

Eckhard, Steffen. 2019. “Comparing How Peace Operations Enable or Restrict the Influence of National Staff: Contestation from Within?” *Cooperation and Conflict* 54(4): 488–505. doi:10.1177/0010836718815528.

Eckhard, Steffen. 2021. “Bridging the Citizen Gap: Bureaucratic Representation and Knowledge Linkage in (International) Public Administration.” *Governance* 34(2): 295–314. doi:10.1111/gove.12494.

Eckhard, Steffen, Vytautas Jankauskas, Elena Leuschner, Ian Burton, Tilman Kerl, and Rita Sevastjanova. 2023. “The Performance of International Organizations: A New Measure and Dataset Based on Computational Text Analysis of Evaluation Reports.” *The Review of International Organizations* 18(4): 753–76. doi:10.1007/s11558-023-09489-1.

Eckhard, Steffen, and Michal Parizek. 2022. “Policy Implementation by International Organizations: A Comparative Analysis of Strengths and Weaknesses of National and International Staff.” *Journal of Comparative Policy Analysis: Research and Practice* 24(3): 254–70. doi:10.1080/13876988.2020.1813032.

Eckhard, Steffen, and Yves Steinebach. 2021. “Staff Recruitment and Geographical Representation in International Organizations.” *International Review of Administrative Sciences* 87(4): 701–17. doi:10.1177/00208523211031379.

Fereday, Jennifer, and Eimear Muir-Cochrane. 2006. “Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development.” *International Journal of Qualitative Methods* 5(1): 80–92. doi:10.1177/160940690600500107.

Ferry, Lauren L., Emilie M. Hafner-Burton, and Christina J. Schneider. 2020. “Catch Me If You Care: International Development Organizations and National Corruption.” *The Review of International Organizations* 15(4): 767–92. doi:10.1007/s11558-019-09371-z.

Ganghof, Steffen. 2005. “Kausale Perspektiven in der vergleichenden Politikwissenschaft: X-zentrierte und Y-zentrierte Forschungsdesigns.” In *Vergleichen in der Politikwissenschaft*, eds. Sabine Kropp and Michael Minkenberg. Wiesbaden: VS Verlag für Sozialwissenschaften, 76–93. doi:10.1007/978-3-322-80441-9_4.

Gutner, Tamar, and Alexander Thompson. 2010. “The Politics of IO Performance: A Framework.” *The Review of International Organizations* 5(3): 227–48. doi:10.1007/s11558-010-9096-z.

Hansen, Henrik, and Finn Tarp. 2000. “Aid Effectiveness Disputed.” *Journal of International Development* 12(3): 375–98. doi:10.1002/(SICI)1099-1328(200004)12:3<375::AID-JID657>3.0.CO;2-M.

Heinrich, Carolyn J. 2012. “Measuring Public Sector Performance and Effectiveness.” *The SAGE handbook of public administration*: 32.

Heinzel, Mirko. 2022a. “Divided Loyalties? The Role of National IO Staff in Aid–Funded Procurement.” *Governance* 35(4): 1183–1203. doi:10.1111/gove.12650.

Heinzel, Mirko. 2022b. “International Bureaucrats and Organizational Performance. Country-Specific Knowledge and Sectoral Knowledge in World Bank Projects.” *International Studies Quarterly* 66(2): sqac013. doi:10.1093/isq/sqac013.

Heinzel, Mirko, Ben Cormier, and Bernhard Reinsberg. 2023. “Earmarked Funding and the Control–Performance Trade-Off in International Development Organizations.” *International Organization* 77(2): 475–95. doi:10.1017/S0020818323000085.

Heinzel, Mirko, and Andrea Liese. 2021. “Managing Performance and Winning Trust: How World Bank Staff Shape Recipient Performance.” *The Review of International Organizations* 16(3): 625–53. doi:10.1007/s11558-021-09414-4.

Hong, Sounman. 2017. “Does Increasing Ethnic Representativeness Reduce Police Misconduct?” *Public Administration Review* 77(2): 195–205. doi:10.1111/puar.12629.

Honig, Dan. 2018. *Navigation by Judgment: Why and When Top-down Management of Foreign Aid Doesn’t Work*. Oxford University Press.

Honig, Dan. 2019. “When Reporting Undermines Performance: The Costs of Politically Constrained Organizational Autonomy in Foreign Aid Implementation.” *International Organization* 73(1): 171–201. doi:10.1017/S002081831800036X.

Honig, Dan. 2020. “Information, Power, and Location: World Bank Staff Decentralization and Aid Project Success.” *Governance* 33(4): 749–69. doi:10.1111/gove.12493.

Honig, Dan, Ranjit Lall, and Bradley C. Parks. 2023. “When Does Transparency Improve Institutional Performance? Evidence from 20,000 Projects in 183 Countries.” *American Journal of Political Science* 67(4): 1096–1116. doi:10.1111/ajps.12698.

ICSC. 2009. *Review of National Professional Officers: terms and conditions of service*. Santiago: International Civil Service Commission (ICSC). https://digitallibrary.un.org/record/675184 (August 5, 2024).

Ika, Lavagnon A. 2015. “Opening the Black Box of Project Management: Does World Bank Project Supervision Influence Project Impact?” *International Journal of Project Management* 33(5): 1111–23. doi:10.1016/j.ijproman.2015.01.005.

Ika, Lavagnon A., Amadou Diallo, and Denis Thuillier. 2012. “Critical Success Factors for World Bank Projects: An Empirical Investigation.” *International Journal of Project Management* 30(1): 105–16. doi:10.1016/j.ijproman.2011.03.005.

Jankauskas, Vytautas, and Steffen Eckhard. 2023. *The Politics of Evaluation in International Organizations*. 1st ed. Oxford University PressOxford. doi:10.1093/oso/9780192855206.001.0001.

Kilby, Christopher. 2015. “Assessing the Impact of World Bank Preparation on Project Outcomes.” *Journal of Development Economics* 115: 111–23. doi:10.1016/j.jdeveco.2015.02.005.

Kullenberg, J. 2016. “Community Liaison Assistants: A Bridge between Peacekeepers and Local Populations.” *Forced Migration Review* (53). https://ora.ox.ac.uk/objects/uuid:fd86ad76-52cf-4c12-a1f4-7eb516168865 (July 22, 2024).

Lall, Ranjit. 2017. “Beyond Institutional Design: Explaining...Ormance of International Organizations.” *International Organization* (2): 245–80.

Lannon, John, and John N. Walsh. 2020. “Project Facilitation as an Active Response to Tensions in International Development Programmes.” *International Journal of Project Management* 38(8): 486–99. doi:10.1016/j.ijproman.2020.06.002.

Malik, Rabia, and Randall W. Stone. 2018. “Corporate Influence in World Bank Lending.” *The Journal of Politics* 80(1): 103–18. doi:10.1086/694102.

Marchesi, Silvia, and Tania Masi. 2021. “Delegation of Implementation in Project Aid.” *The Review of International Organizations* 16(3): 655–87. doi:10.1007/s11558-020-09396-9.

Mosley, Paul. 1986. “Aid-Effectiveness: The Micro-Macro Paradox.” *IDS Bulletin* 17(2): 22–27. doi:10.1111/j.1759-5436.1986.mp17002004.x.

Ogwezzy, Michael. 2016. “The Noblemaire Principle in the Context of International Civil Service.” *Cosmopolitan Law Journal* 4(2): 14–28.

Olken, Benjamin A. 2007. “Monitoring Corruption: Evidence from a Field Experiment in Indonesia.” *journal of political economy*.

Panke, Diana. 2021. “Compensating for Limitations in Domestic Output Performance? Member State Delegation of Policy Competencies to Regional International Organizations.” *International Relations* 35(1): 90–125. doi:10.1177/0047117820970320.

Parizek, Michal. 2017. “Control, Soft Information, and the Politics of International Organizations Staffing.” *The Review of International Organizations* 12(4): 559–83. doi:10.1007/s11558-016-9252-1.

Parizek, Michal. 2024. “Less in the West: The Tangibility of International Organizations and Their Media Visibility around the World.” *The Review of International Organizations*. doi:10.1007/s11558-024-09551-6.

Parizek, Michal, and Matthew D Stephen. 2021. “The Increasing Representativeness of International Organizations’ Secretariats: Evidence from the United Nations System, 1997–2015.” *International Studies Quarterly* 65(1): 197–209. doi:10.1093/isq/sqaa088.

Paterson, Ted, and Vera Bohle. 2008. *Evaluation of the UNDP Sudan Mine Action Capacity Building and Development Project*. Geneva: UNDP.

Platteau, Jean-Philippe. 2004. “Monitoring Elite Capture in Community-Driven Development.” *Development and Change* 35(2): 223–46. doi:10.1111/j.1467-7660.2004.00350.x.

Raimondo, Estelle. 2016. “What Difference Does Good Monitoring and Evaluation Make to World Bank Project Performance?”

Raimondo, Estelle. 2018. “The Power and Dysfunctions of Evaluation Systems in International Organizations.” *Evaluation* 24(1): 26–41. doi:10.1177/1356389017749068.

Reinsberg, Bernhard, Andreas Kern, Mirko Heinzel, and Saliha Metinsoy. 2024. “Women’s Leadership and the Gendered Consequences of Austerity in the Public Sector: Evidence from IMF Programs.” *Governance* 37(1): 303–21. doi:10.1111/gove.12764.

Saldana, Johnny. 2015. “The Coding Manual for Qualitative Researchers.” : 1–440.

Schoonenboom, Judith, and R. Burke Johnson. 2017. “How to Construct a Mixed Methods Research Design.” *Kolner Zeitschrift Fur Soziologie Und Sozialpsychologie* 69(Suppl 2): 107–31. doi:10.1007/s11577-017-0454-1.

Seawright, Jason, and John Gerring. 2008. “Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options.” *Political Research Quarterly* 61(2): 294–308. doi:10.1177/1065912907313077.

Shin, Wonkyu, Youngwan Kim, and Hyuk-Sang Sohn. 2017. “Do Different Implementing Partnerships Lead to Different Project Outcomes? Evidence from the World Bank Project-Level Evaluation Data.” *World Development* 95: 268–84. doi:10.1016/j.worlddev.2017.02.033.

Sommerer, Thomas, Theresa Squatrito, Jonas Tallberg, and Magnus Lundgren. 2022. “Decision-Making in International Organizations: Institutional Design and Performance.” *The Review of International Organizations* 17(4): 815–45. doi:10.1007/s11558-021-09445-x.

Sundberg, Molly. 2019. “Donors Dealing with ‘Aid Effectiveness’ Inconsistencies: National Staff in Foreign Aid Agencies in Tanzania.” *Journal of Eastern African Studies* 13(3): 445–64. doi:10.1080/17531055.2019.1628384.

Sundberg, Molly. 2023. “Local Recruits in Development Finance Institutions: Relocating Global North-South Divides in the International Aid Industry.” *The Journal of Development Studies* 59(11): 1635–51. doi:10.1080/00220388.2023.2232915.

Takasaki, Yoshito. 2011. “Do Local Elites Capture Natural Disaster Reconstruction Funds?” *Journal of Development Studies* 47(9): 1281–98. doi:10.1080/00220388.2010.509786.

Tallberg, Jonas, Thomas Sommerer, Theresa Squatrito, and Magnus Lundgren. 2016. “The Performance of International Organizations: A Policy Output Approach.” *Journal of European Public Policy* 23(7): 1077–96. doi:10.1080/13501763.2016.1162834.

Tammi, Iida-Maria. 2022. “The Humanitarian Frame of War: How Security and Violence Are Allocated in Contemporary Aid Delivery.” *Third World Quarterly* 43(5): 963–78. doi:10.1080/01436597.2022.2055538.

Tashakkori, Abbas, and Charles Teddlie. 2003. *Handbook of Mixed Methods in Social & Behavioral Research*. Thousand Oaks, Calif.: SAGE Publications.

UN WOMEN. *Personnel Categories and Basic Requirements*. UN WOMEN. https://www.unwomen.org/sites/default/files/Headquarters/Attachments/Sections/About%20Us/Employment/UN-Women-personnel-categories-en.pdf (August 28, 2024).

UNCEB. 2023. “UN System HR Statistics Report - 2022.” https://unsceb.org/un-system-hr-statistics-report-2022 (July 25, 2024).

UNDP. 2024. “Staff.” *UNDP*. https://www.undp.org/careers/types-opportunities/staff (July 25, 2024).

UNEG. 2016. “Norms and Norms and Standards for Evaluation.”

UNICEF. 2024. “UNICEF Job Categories | UNICEF Careers.” https://www.unicef.org/careers/unicef-job-categories (July 25, 2024).

Ward, Patricia. 2021. “Capitalising on ‘Local Knowledge’: The Labour Practices behind Successful Aid Projects – the Case of Jordan.” *Current Sociology* 69(5): 705–22. doi:10.1177/0011392120905342.

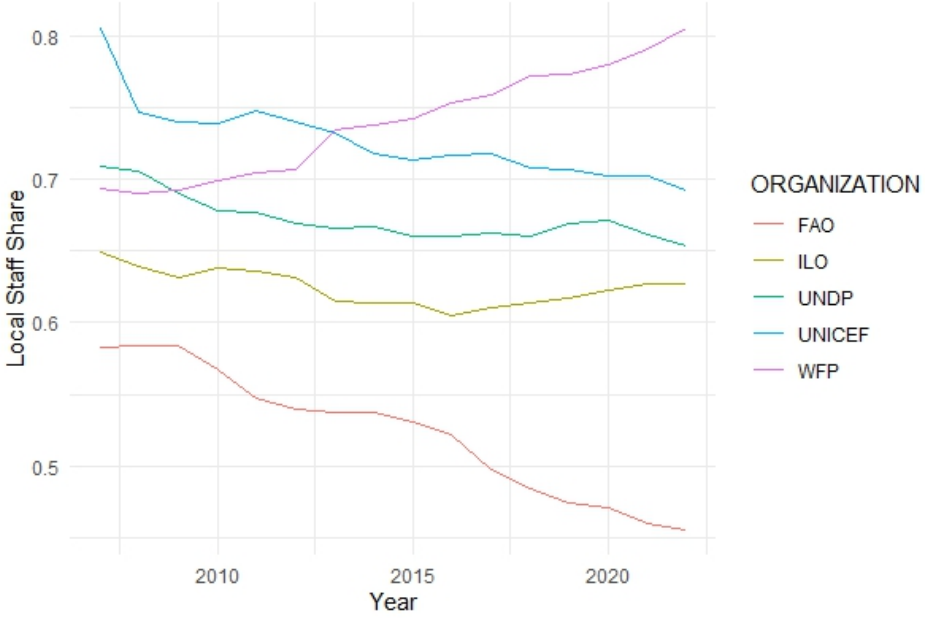
Weaver, Catherine, Mirko Heinzel, Samantha Jorgensen, and Joseph Flores. 2022. “Bureaucratic Representation in the IMF and the World Bank.” *Global Perspectives* 3(1): 39684. doi:10.1525/gp.2022.39684.

Winters, Matthew S. 2014. “Targeting, Accountability and Capture in Development Projects.” *International Studies Quarterly* 58(2): 393–404. doi:10.1111/isqu.12075.

Winters, Matthew S. 2019. 2019 *Too Many Cooks in the Kitchen? The Division of Financing in World Bank Projects and Project Performance*. 6th ed. UNU-WIDER. doi:10.35188/UNU-WIDER/2019/640-1.

APPENDIX

Appendix I: Local Staff Share of Selected IOs



Appendix II: Lebanese Projects in Dataset

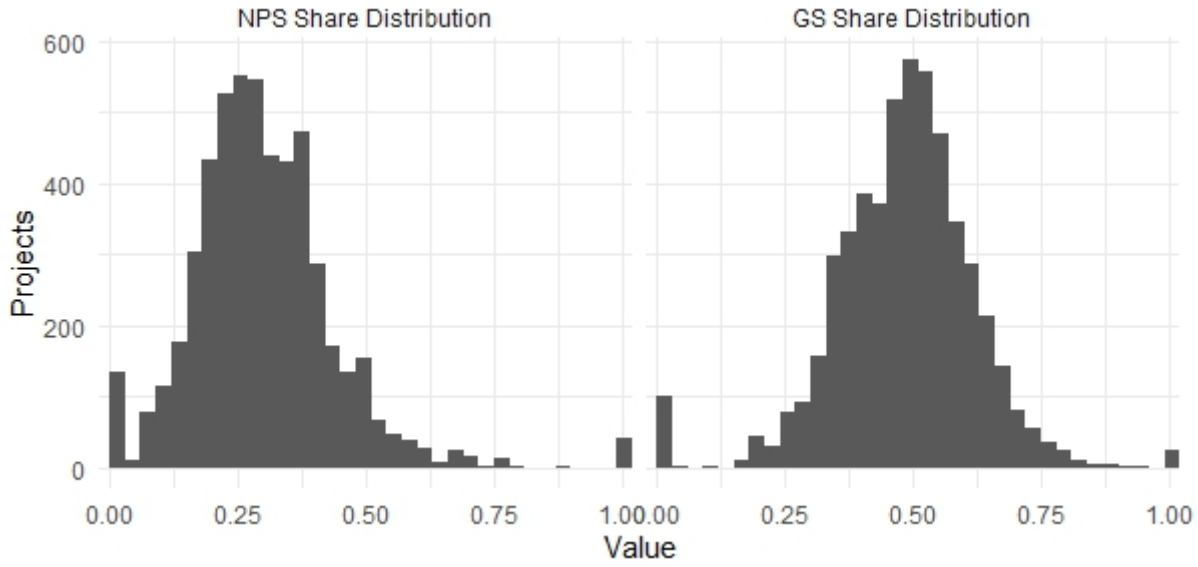
	Total Dataset		*Lebanon*
Variable	*Mean*	*SD*	*Mean*
Performance (0-1)	0.567	0.130	0.562
LS share (0-1)	0.787	0.134	0.738
NPO share (0-1)	0.301	0.138	0.255
GS share (0-1)	0.486	0.133	0.483
Report Length (sentences)	1222.32	716.99	930.20
Joint Evaluation (binary)	0.095	0.294	0.014
Mid-Term Evaluation (binary)	0.151	0.358	0.086
Number of Offices	2.29	2.54	1.77
Liberal Democracy (0-1)	0.307	0.196	0.282
State Fragility (0-120)	81.20	15.01	88.35
GDP per Capita (current USD)	4413.34	7063.78	7828.00
Population (log)	16.70	1.86	15.55
ODA Total (log)	20.30	1.31	20.61
N (Projects)	5,176		70

Appendix III: Overview of included IOs

IO	*Acronym*	*Budget* (bn USD)*	*Total Staff**	*Total Offices**	*Projects covered*
Food and Agriculture Organization of the United Nat.	FAO	2.854	3202	159	111
International Labour Organization	ILO	0.849	3651	183	327
United Nations Development Programme	UNDP	5.284	7394	285	3520
United Nations International Children's Emergency Fund	UNICEF	10.329	15654	413	1105
World Food Programme	WFP	14.418	11670	411	113
Total	-	33.734	41,571	-	5176
Share of UN system overall	-	0.454	0.331	-	-

*Year of reference is 2022

Appendix IV: Distribution of Local Staff Categories



Appendix V: Local Staff Share and Performance Correlations by IO