Assessing the Evaluation Methods of the Millennium Challenge Corporation's Latin American Compact Projects

Public Policy and Romance Studies Honors Thesis

Savannah Norman

Thesis Advisor: Dean Judith Kelley, Ph.D.

Section Advisor: Professor Christina Gibson-Davis, Ph.D.

Romance Department Independent Study Lead: Professor Gustavo Furtado, Ph.D.

20 November 2020

Acknowledgements

I would like to sincerely thank the Duke University faculty and administration members who advised and molded my thesis. First, Dean Judith Kelley, who shaped the current investigation, provided keen insights, and never failed to ask, “So, how are you doing?” Second, Dr. Gustavo Furtado, who always pushed me to consider different perspectives, encouraged me to look beyond surface level, and motivated me to use my voice. Next, Professors Esther Gabara and Walter Mignolo for, along with Dr. Furtado, serving on my thesis defense committee. Finally, Dr. Christina Gibson-Davis, who taught me how to write a thesis in the first place, answered my many questions, and provided the feedback necessary to bring this paper to its present level.

I am incredibly grateful to all of my loved ones – family, significant other, and friends – for their continual support and encouragement throughout this year-long process.

Lastly, I am grateful to the authors of the evaluations who graciously participated in candid interviews. My analysis benefitted greatly from their insights. On a larger scale, however, I am grateful for their commitment to the field of development – their work is meaningful and has the capacity to inspire true learning and change.

Table of Contents

Abstract

International development agencies have, for decades, worked to remedy global development challenges. One of these agencies is the Millennium Challenge Corporation (MCC). This study explores the MCC's Latin American programming via its evaluations. It specifically assesses the technical soundness of the MCC's Honduras, Nicaragua, and El Salvador evaluations, as well what the findings and soundness of evaluations mean for future projects in the region. Through document analysis, this study found that the average evaluation was technically sound, as it employed consistent, appropriate, and objective metrics. However, evaluation structure varied according to author and proprietor agency. External evaluations did not directly contradict the findings of MCC-sponsored evaluations. Evaluations were not inclusive to the intended recipients of MCC programming, beneficiary groups. Interviews with the authors of evaluations confirmed these findings. The implications of findings include the importance of culturally competent, inclusive, and multi-faceted development processes that assimilate learning from prior programming. Findings are applicable to development and evaluation processes in Latin America.

Introduction

International development initiatives, which can be defined as "all social and economic programs in developing countries funded by…development agencies" began in the years following World War II (Bamberger 2000, pg. 96; Black 2015). These initiatives were envisioned to be a process through which developing nations could "catch up" with the industrialized, 'developed' nations of the global North. Since its origins, international development has ballooned into a huge industry, providing billions of dollars to Africa, Asia, the Middle East, parts of Europe and Latin America so that they might develop economically and scale their economies. This economic development was envisioned to "trickle down," bringing social, educational, and health improvements along with it (Mowforth 2014). Despite the billions expended, development initiatives are not always successful and often fail to meaningfully affect the lives of impoverished individuals (Black 2015). Such variability of impact in the development process necessitates a thorough evaluation of development programming, so that future initiatives can grow from previous failures and improve in their provision of lasting impact. These evaluations are the subject of the present study, specifically those of the Millennium Challenge Corporation.

The Millennium Challenge Corporation (MCC) is a United States government agency that facilitates international development initiatives in developing countries. Since its establishment in 2004, the MCC has implemented 37 compact projects (large-scale, 5-year agreements with recipient nations) in 29 countries that constitute annual government expenditures of $900 million. The MCC operates under a conditional aid scheme, meaning that it provides economic aid through a competitive selection process. To do so, the MCC selects developing nations that demonstrate positive performance in three sectors – ruling justly, investing in people, and fostering economic freedom. Once the MCC selects countries that have performed well according to these conditions, the agency enters into the planning phase of each project and then implements the project over a five-year timeframe. The ultimate goal of these projects is to reduce poverty through economic growth (Brown 2019).

After the implementation phase, the MCC contracts independent agencies to evaluate completed projects. These evaluations are purported to measure either the impact projects had on beneficiaries or the extent to which the project achieved its goals. The purpose of this paper is to assess these evaluations. The evaluations I study stem from the MCC's Latin American projects – its El Salvador, Honduras, and Nicaragua compacts. My analysis seeks to answer two questions: To what extent are the evaluations of the Millennium Challenge Corporation's Latin American compact projects of high quality? And, what do the quality and findings of these evaluations mean for the MCC's ongoing and future projects in the region?

Recent research has examined the general evaluation practices of the MCC and the difficulty associated with actually evaluating the impact of these projects (BenYishay and Tunstall 2011, Sturdy, Aquino et al. 2014, Brown 2019). However, little work has been done to examine these evaluations with a regional lens nor determine how future projects can move forward with greater success in the region. As such, recommendations to improve development in the region should have a uniquely Latin American character, so that they can solve uniquely Latin American problems.

My findings contribute to scholarship related to international development initiatives and evaluation practices in Latin America in several ways. First, by analyzing the MCC's Latin American evaluation reports, I reveal the extent to which evaluation efforts are being executed in a quality manner. I also reveal that which is missing from these reports – namely the consideration of project impact on beneficiaries. Second, by interviewing those responsible for conducting these evaluations, I am able to gauge the extent to which regional considerations and beneficiary impact are important to the evaluations process. Finally, from my data collection, I am able to make recommendations for how the MCC can improve both its project structure and implementation and evaluation efforts in the region.

This research is important because the MCC is currently implementing one project in Latin America and recently terminated another. My research can have significant implications for the remaining phases of these projects, including the evaluation phase. Moreover, research such as this is needed because US foreign assistance in the region constitutes at least $1 billion in annual expenditures (Meyers and Garcia 2019). This figure represents a cut to Latin American funding, as the Trump administration continues to scale back U.S. assistance to the region (Meyers and Garcia 2019). Regardless, given the size of expenditures, it is important to ensure that these initiatives are being implemented and evaluated in a manner conducive to the MCC's ultimate goal – poverty reduction through economic growth.

Research Questions

To what extent are the evaluations of the Millennium Challenge Corporation's Latin American compact projects of high quality? What do the quality and findings of these evaluations mean for the MCC's ongoing and future projects in the region?

Background Information

The Millennium Challenge Corporation – Establishment & Purpose

The MCC is a United States government agency that operates in the international development field. The MCC was created in 2004 by the Bush administration to address existing frustrations with other international development agencies, namely that "traditionally-implemented" projects had little measurable impact (Brown 2019). The MCC attempted to remedy this frustration with a structural innovation – the MCC requires that potential recipient countries demonstrate their commitment to improvement before programs are implemented or grants are rewarded. This approach – called a conditional aid approach – operates under the assumption that countries that are already making improvements are likely to make even more improvements given the implementation of an MCC program (Öhler, Nunnenkamp et al. 2012). As a US government agency, the MCC receives annual funding from Congress. Although the MCC was initially envisioned to operate with a $5 billion annual budget, Congress has, since 2011, apportioned the MCC a much smaller budget of $900 million annually (Brown 2019).

How does the MCC Function?

Since its establishment in 2004, the MCC has signed 37 grant agreements with 29 countries (Brown 2019). These grants – called compacts – are 5-year, large-scale investments that represent a partnership between the MCC and the recipient country. As the conditional aid approach is central to the MCC's strategy, the country selection process is competitive. The process itself has three steps – candidate country identification, determination of selection criteria and methodology, and country selection for compact eligibility. The first step is determined by two indicators weighing the financial need and legal status of potential recipient countries. To be eligible, countries must be classified by the World Bank as low-income or lower-middle income nations and must fall within the 75 lowest per capita income countries for that fiscal year. Secondly, potential recipient countries must be free of legal prohibitions on foreign aid (Brown 2019). For a visual representation of the aforementioned requirements, see Table 1.

Once candidate countries are identified, the MCC creates selection criteria against which to evaluate these countries. This selection criteria serves to represent the MCC's development priorities and constitutes an important part of the MCC's conditional approach to aid. Broadly defined, these priorities can be separated into three categories: ruling justly, investing in people, and fostering economic freedom. The MCC defines ruling justly as the extent to which a country promotes democratic institutions, fights corruption, and protects basic human rights. Investing in people has to do with the recipient country's provision of necessary resources for its citizens such as health care and educational services. Finally, the MCC defines fostering economic freedom as a country's commitment to neoliberal practices such as free market trade and globalized economic practices (Brown 2019). In other words, potential recipient countries must make improvements in and show commitment to these three criteria before ever receiving MCC funding – the crux of the conditional aid approach. The provision of MCC aid dollars is conditional on recipient countries' commitment to and action in the categories of ruling justly, investing in people, and fostering economic freedom. For the official definitions of these three criteria, see Table 2.

The final step consists of evaluating candidate countries against these criteria. To do so, the MCC quantifies the three categories, converting them into 20 performance indicators. Candidate countries are evaluated according to their commitment to these indicators and are scored based on their respective progress-based performances. The MCC then evaluates the scores of each candidate country, examining each country against others of its same income group and regional context. Countries that "pass" this examination typically score higher than the median of its peer countries. Passing, however, does not guarantee grant eligibility, nor does "failing" preclude eligibility. The MCC Board often considers additional data before making final eligibility decisions (Brown 2019).[1]

Furthermore, the MCC reserves the right to terminate compacts at any point in their five-year lifespans if the recipient government backslides in the categories of ruling justly, investing in people, or fostering economic freedoms. Such was the case for the Nicaragua compact, which was partially terminated in 2009 following the Nicaraguan government's perpetration of fraudulent municipal elections (MCC, *Congressional Notification*). As a result, the Nicaragua *Transportation* and *Property Regularization* sub-activities[2] were ended prematurely.

In Latin America specifically, The MCC has implemented very few projects – only 10 of its 79 total projects are located in the region (MCC, *Where We Work*). Three of these ten projects are compact projects – those which I will study – while the remaining seven are threshold projects. (Threshold projects are smaller-scale, shorter-term projects). In total, the MCC's ten completed projects in Latin America constitute an expenditure of $902,331,092 over the compact lifetimes. Further, the MCC is in the implementation stage for one project in Latin America, the Guatemala Threshold Program (2015), and recently completed another, the El Salvador

---

[1] Note: The countries whose compacts I will be studying – El Salvador, Honduras, and Nicaragua – underwent this three-step process and were ultimately selected for compact development and implementation.

[2] Note: Every italicized project name represents one of the sub-activities of the Honduras, Nicaragua, or El Salvador compacts. Within the Honduras compact, there are three sub-activities. Within the Nicaragua compact, there are six sub-activities. Finally, within the El Salvador compact, there are 9 sub-activities. Each sub-activity has its own evaluation document, which comprise the 18 evaluations included in the document analysis. For a complete list of the sub-activities, see Table 4.

Investment Compact (2014), in September 2020. Together, these projects constitute an expenditure of $305,000,000 (MCC, *Where We Work*).

Foreign Aid in Latin America – Brief History & Critiques

The United States has long contributed foreign aid dollars to Latin America, with $181 billion expended from 1946 to 2017. Those supportive of former and current aid initiatives in the region see US humanitarian and aid dollars as beneficial to developing nations. In the late 1990s and early 2000s particularly, the US contributed significantly to post-disaster relief initiatives as well as public health campaigns (Meyer and Gracia 2019).

However, US foreign aid initiatives are also heavily criticized for doing more harm than good and for serving often-dubious political and economic interests (Harvey 2007, Mowforth 2014). The World Bank and the International Monetary Fund – in their efforts to spark economic growth in Latin American nations through market reforms – are instead criticized for indebting recipient nations and deepening class divides (Harvey 2007). US assistance dollars meant to enhance the 'rule of law' in Central America are instead criticized for training and equipping murderous regimes and paramilitary units – from the Contras in Nicaragua in the 1980s to those responsible for the Honduran military coup in 2009 (Mowforth 2014). Finally, and in recent years, US aid initiatives in Latin America have been criticized as exploitative of natural resources and of the communities of indigenous people and rural peasants that depend on these resources (Lang and Mokrani 2013).

The harshest of critics describe US foreign aid initiatives as inherently violent processes by which developing nations – and their most vulnerable citizens – are exploited to line the pockets of large corporations. Others question the intentions of US aid initiatives, claiming that they are implemented to serve US strategic interests and all but neglect beneficiary populations (Mowforth 2014). The importance of US strategic interest in foreign aid is even clearly stated in a US Congressional Research Service report – the US provides aid dollars to Latin America to "support development and other US objectives" (Meyer 2018, pg. ii). Several Latin American nations seem to share these skeptical views of US foreign aid initiatives, as both Bolivia and Ecuador expelled US development professionals, refused US-sponsored projects, and closed USAID offices in recent years (Meyer and Gracia 2019).

Following these developments in Ecuador and Bolivia, as well as independent economic growth in other large South American nations, US aid dollars in recent years have been directed toward Central American nations (Meyer and Gracia 2019). This shift is well-represented by the MCC's Latin American initiatives, as all of its compact projects are located in Central America. A final consideration for foreign aid in the region is the Trump administration's attempted rollbacks on funding and programming. When planning the 2019 fiscal year, for example, the Trump administration proposed a 34% slash to funding in Latin America (Meyer and Gracia 2019).

Literature Review

The goal of this literature review is to identify both the general evaluation practices of international development agencies as well as the specific evaluation practices of the MCC. This literature review will also outline common criticisms of these evaluation practices and ways in which they can be improved. The main criticism of these evaluations is that they are donor-centric rather than beneficiary-centric, meaning that they are written to satisfy the needs of project donors rather than to assess the extent to which projects actually aided their intended beneficiaries. Specific criticisms of the MCC's evaluations are similar but also emphasize the MCC's use of questionable data and tendency to overstate impacts. Remedies to these criticisms are that evaluations be conducted in a manner inclusive to project beneficiaries and conducive to learning and improvement. This literature review also identifies the characteristics of evaluations that are both of quality and are useful. Such evaluations have a well-defined scope, accurate data, sound analysis, objective findings and conclusions, clear wording, and must actively cater to the audiences best situated to act upon the contents of the evaluation. Finally, this literature review explains and motivates the focus on Latin America by expanding upon the development challenges unique to the region. These challenges include the mass inequalities of the region, its diverse ethnic and cultural composition, its history of authoritarian regimes and political instability, and its current and historic struggles with violence and corruption. Such unique challenges demonstrate the importance of cultural context in the development process.

Project Evaluations in International Development – What do we know?

Most international development agencies require the evaluation of their programs (Bamberger 2000). Often, these evaluations are conducted by independent organizations, as is the case for the MCC (Brown 2019). Although the term "evaluation" is used differently by different agencies, many of these distinguish between monitoring activities and evaluation activities. Monitoring typically occurs during the implementation phase of a project and seeks to measure the extent to which the project is progressing according to plan. Evaluation activities, however, occur after a project has been completed and seek to measure the extent to which projects either fulfilled their intended goals or enacted change in recipient communities (Bamberger 2000).

The usefulness of current evaluation methods in the international development field is questioned. At a minimum, monitoring and evaluation activities are useful to international development agencies in their fundraising efforts. When these agencies are able to demonstrate quantifiable results and measurable impacts to donors, donors are more likely to continue to fund these projects. Projects without measurable results or with ostensibly-weak impacts, on the other hand, are likely to receive less funding (Hayes, Introna et al. 2018, Brown 2019). International development agencies, consequently, have adapted to fit this results-based format. Such is the case for the MCC, which heavily values its ability to demonstrate measurable results for each of its projects (Sturdy, Aquino et al. 2014). This results-based approach is sometimes criticized as donor-centric, meaning that evaluations are designed to please and privilege donor entities. (Hayes, Introna et. al 2018; Chouinard and Hopson 2016). Further criticisms suggest that evaluations in international development were "borne out of the need of funding agencies," designed only to meet donor requirements (Chouinard and Hopson 2016, pg. 250). As a result of this alleged donor-centricity, the "real" question of impact – the extent to which programs actually benefitted its intended recipients – goes unanswered in these evaluations. Further, learning and improvement opportunities are often forgone (Guerra-López and Hicks 2015). These critcisms extend beyond the theoretical; a meta-analysis of 71 evaluation documents found most to prioritize donor accountability purposes over the portrayal of project impact on beneficairy communities (Chouinard and Hopson 2016).

Despite these criticisms, evaluations still have the potential to improve. If evaluations are conducted in a manner conducive to learning, improvement, and inclusivity, they have the capacity to examine both the effectiveness of development interventions and their impacts on beneficiaries (Hayes, Introna et al. 2018; Peters 2016).

To remedy concerns surrounding donor-centricity and the subsequent exclusion of project beneficiaries in the evaluations process, evaluators have designed several methods. The first of these are participatory and thematic evaluation methods (Bamberger 2000; Guerra-López and Hicks 2015). Participatory evaluations use qualitative methods to give voice to vulnearble stakeholders and intended project beneficiaries in both the project implementation and evaluation phases. Thematic evaluations take this a step further, examining the effects of an international development intervention from the perspective of a specific group, such as indigenous peoples or women (Bamberger 2000). These methods can serve to answer the question of project impact by directly seeking the opinions and experiences of those affected by the project. Ideally, these qualitative methods are meant to complement and be integrated with other quantitative evaluation data. Another method to promote inclusivity in the evaluation process is publishing evaluations in the native language(s) of recipient countries (Hayes, Introna et al. 2018). Finally, it is recommended that evaluations move to promote an approach of learning and improvement rather than a mere display of project results (Archibald, Sharrock et al. 2018). From such an approach of learning, future projects can capitalize on insights from previous evaluations, incorporating methods to foster inclusivity in both the implementation and evaluation phases.

What makes a quality evaluation?

In addition to these concerns on the goals and inclusionary character of evaluations, it is also important to explore the structural making of a quality evaluation. Four criteria contribute to evaluation quality – a well-defined scope, accurate data, sound analysis, and objective findings and conclusions (Schwartz and Mayne 2005). First, a well-defined scope, which clearly defines the objectives, purposes, and range of evaluations. Second, accurate data, which provides evaluations with validity, reliability, and replicability. Next, sound analysis, which ensures that data are analyzed with robust methodologies. Finally, objective findings and conclusions ensure that these findings and conclusions are supported with reliable data and are presented objectively. For complete definitions of these criteria, see Table 3.

However, quality evaluations should also be useful, providing insights and facilitating future learning. For an evaluation to be useful, the evaluator must consider her audience(s) and what these audience(s) needs from the evaluation (Grasso 2003). Regarding the first consideration, a useful evaluation must cater to the primary audience of the evaluation – likely the organization that commissioned it – as well as those audiences best positioned to do something with the contents of the evaluation. In the context of international development, this audience is likely the managers and implementors of the evaluated projects. If evaluation contents are suited to this audience, the evaluation can impart learning and facilitate structural improvements for future projects (Néstor Feinstein 2012). Finally, to create a useful evaluation, the evaluator must fulfill the needs of her audiences. To do so, she must provide pertinent information in an easily understandable, clear, and effective manner (Grasso 2003).

How does the MCC Compare: A look into its evaluation methods

The MCC highly values the ability to demonstrate measurable results from its projects. As such, it strictly adheres to a result-based approach and measures these results in all phases of the compact – from compact development, to implementation, to completion (Brown 2019). In the development stage, the MCC predicts economic rates of return for its projects, or the anticipated net economic benefits that project beneficiaries will receive once a project has been implemented (Brown 2019). In the implementation phase, the MCC monitors the ongoing results of its projects. To do so, it appoints an MCC staff member to each project to oversee monitoring and evaluation (M&E) efforts and creates an entire M&E unit for each project. To actually monitor compact progress, this unit collects quarterly data and compares it to the development-phase predictions (Brown 2019, Sturdy, Aquino et al. 2014). After compact completion, the MCC contracts external, independent agencies or individuals to conduct evaluations. These agencies or individuals report to have complete freedom regarding evaluation design, implementation, and findings (Sturdy, Aquino et al. 2014). Once complete, the MCC publishes these evaluations on its website, regardless of the findings.

The MCC distinguishes two evaluation categories – impact evaluations and performance evaluations – along with a third category that combines the two, called “multiple.” Of the two categorizations, impact evaluations are the more costly, as they seek to measure changes directly attributable to MCC interventions. In other words, these evaluations estimate a counterfactual, depicting what would have happened to project beneficiaries had the project not occurred (Sturdy, Aquino et al. 2014). Performance evaluations, on the other hand, measure the extent to which a project completed its goals and remained true to its originally envisioned purposes (Brown 2019). As of August 2019, the MCC has published 32 impact evaluations and 51 performance evaluations. For its Latin American projects specifically, the MCC has published eight performance evaluations, six impact evaluations, and four multiple-method evaluations (MCC, Independent Evaluations).

Finally, the MCC uses multiple independent agencies and private individuals to evaluate its compacts. For its Latin American compact projects alone, the MCC contracted three separate agencies and six private individuals (MCC, Independent Evaluations). Given the variance in evaluators, the evaluations themselves vary in structure, design and methods.

What are the criticisms of the MCC's evaluation efforts?

Given the MCC's results-based approach and desire to show concrete findings from its projects, its first wave of evaluations were primarily impact evaluations – those that seek to establish counterfactuals. However, these early impact evaluations faced many problems and were criticized by outside entities such as the Government Accountability Office (GAO). The primary problems that these early evaluations faced were: a disconnect between program design and implementation and evaluation design; a failure to assess impacts along the causal chain; and a failure to allot sufficient time to evaluations efforts. The result of these issues was an overall difficulty in establishing sound counterfactuals to the MCC projects and a consequential failure to accurately demonstrate the effects of MCC interventions on intended beneficiaries (Sturdy, Aquino et al. 2014).

Since discovering the difficulty associated with producing sound impact evaluations, the MCC claims to have improved its program design and impact measurement practices (Brown 2019). The main strategy to do so is an increased selectivity in using impact evaluations versus performance evaluations (Sturdy, Aquino et al. 2014). Such an increased selectivity in producing impact evaluations means an increased reliance on performance evaluations, which seek to examine the extent to which projects achieved their original goals (Archibald, Sharrock et al. 2018, Brown 2019). Although the MCC has claimed to strenghten its performance evalutions by designing them as more conducive to learning, they cannot show direct project impact on beneficiaries (Sturdy, Aquino et al. 2014).

Why Latin America?

As described in the previous section "*Foreign Aid in Latin America – Brief History & Critiques*," the United States has had a long and complicated presence in Latin America. Indeed, the US's propagation of neoliberalism in the region is motivation enough to study development with a Latin American lens (Blandón 2001). However, there are other factors that motivate the regional focus of the present study, including the region's inequality, ethno-cultural diversity, and history with authoritarianism, political instability, and violence. The interplay of these factors make difficult the extrapolation of study from one region to another; it would be reductionist to assume that all developing regions face the same problems, and that these problems can be solved with the same solutions.[3] As such, development initiatives – and their subsequent evaluations – should be examined in specific, regional contexts (Romero 2005).

The first of these factors is societal inequality. Although inequality is prevalent in both developed and developing regions, inequality in Latin America has been recognized as the most extreme since the 1980s (UNDP 2019). This inequality manifests not only in income disparities but also in disparities in access to services, such as healthcare, education, and sanitation (Flores et. al 2013; Rivero 2000; Ramírez 2014; Kliksberg 1999). In Honduras, Nicaragua, and El Salvador specifically, poverty proliferates, with 75.1% of Hondurans, 64.6% of Nicaraguans, and 42.1% of Salvadorans living in conditions of poverty or extreme poverty (CEPAL 2019). This poverty is also accompanied by inequality. According to the Gini Coefficient, which represents inequality along a scale of zero (perfect equality) to one (perfect inequality), Honduras and Nicaragua fall into the second-highest category of inequality. Although El Salvador fell into the

[3] Note: Ironically, and as I will discuss in the Results and Discussion sections, the MCC adheres to this reductionist approach to development.

lowest Latin American category of inequality (receiving between a 0.378 and 0.418), this classification aligns it with present-day Venezuela.

Unfortunately, such inequality often disproportionately affects ethnic minority groups, such as indigenous and afro-descendent populations. These inequalities can be traced back to the colonial period, during which the Spanish and Portuguese enslaved indigenous peoples and stolen Africans (Quijano 2000; Flórez, Medina et. al 2003; Deruyttere 1997; Telles and Flores 2013). The power dynamics of colonialism have persisted – the wealthiest of Latin Americans are often direct descendants of European colonizers and oligarchs. The most impoverished and disenfranchised groups, conversely, are indigenous and afro-descended (Flórez, Medina et. al 2003; Deruyttere 1997; Costa 2012). Honduras, Nicaragua, and El Salvador each have significant – although minority – populations of indigenous peoples, with these groups comprising 15%, 5%, and 7% of the national populations, respectively. Afro-descendent populations are much smaller, comprising 1.4% of the Honduran population, 2.8% of the Nicaraguan, and 0.1% of the Salvadoran. Regardless of population size, both of these groups are more likely to live in conditions of poverty than are their non-minority counterparts (Banco Mundial 2018, Deruyttere 1997).

The ethnic diversity of Latin America – independent of its relation to socioeconomic inequalities – motivates the regional examination of Latin America. For example, culture and shared values may vary significantly by ethnic group; indigenous peoples and certain afro-descendant populations may practice distinct religions or hold non-Western world views (Romero 2005). Certain indigenous groups, specifically, may adhere to non-accumulative conceptions of wellbeing or may value services other than access to formal education or healthcare (Deruyttere 1997).

Other factors important to the isolated study of Latin America are the region's histories of authoritarian regimes, political instability and violence. These factors – and especially those of instability and authoritarianism – have plagued almost every Latin American nation (Díaz, 2014; Ducatenzeiler, 1994; Winton 2009; Blanco 2008). Honduras, Nicaragua, and El Salvador are no exceptions, having experienced both historical and present-day instability and borderline authoritarianism (García-Pinzon and Rojas Ospina 2020; Chaguaceda 2015; Herrera 2018). The twentieth century for these three nations was particularly marked by military dictatorships, and present conditions are not much brighter. In 2009, for example, the Honduran government was overthrown via a military coup d'état (Herrera 2018). The democratic conditions of Nicaragua have rapidly deteriorated following a series of fraudulent elections, beginning in 2008, and the violent repression of civilian protests in the spring of 2018 (MCC, Congressional Notification 2009; GIEI 2018). And, El Salvador's President, Nayib Bukele, continually slides toward militant authoritarianism (García-Pinzon and Rojas Ospina 2020). Conditions of authoritarianism and instability are not only threatening to democratic processes but have also been shown to perpetuate societal inequalities (López 2018). What's more, the violence associated with the traffic of narcotics both contributes to existing levels of instability and encourages mass exoduses in the form of emigration (Winton 2009; Jiménez 2016).

Although the above-described factors are by no means an exhaustive list of the characteristics unique to Latin America and important to its discourse of development, they regardless serve to motivate the present investigation and justify its regional focus.

What can we learn from project evaluations & why do we care?

Given the questionable usefulness, quality, and inclusivity of evaluations in the international development field, the act of assessing the MCC's Latin American compact evaluations can have important implications for future and ongoing projects – and their respective evaluation practices – in the region. An analysis of evaluations can show whether modifying evaluation practices, such as reducing variation in evaluators and standardizing evaluation structure, can improve evaluation quality. This analysis can also show whether the MCC is learning from previous projects and applying this learning to future projects. Finally, this analysis can show whether MCC-sponsored evaluations are donor-centric and fail to meaningfully consider project beneficiaries.

If evaluations do fall short in their inclusion of beneficiaries, my analysis can have important recommendations for evaluation improvement. Especially because the evaluations included in the analysis share a common geographic space, these recommendations can be regionally specific and suggest best practices for a culturally competent development process. These potential insights are particularly important given the ongoing nature of MCC implementation efforts for the Honduras Threshold Program and imminent evaluation efforts for the El Salvador Investment Compact.

Regardless of my findings, the act of assessing these evaluations is important because the MCC annually spends hundreds of millions of dollars globally, and tens of millions of dollars in Latin America. It is important to ensure that these multi-million dollars projects are being evaluated effectively and are assessing the extent to which the MCC has achieved its primary goal – poverty reduction in beneficiary populations (Brown 2019). Subsequent to this point is that, if individuals have confidence in the evaluations, they will also have confidence in the evaluation findings, which could then inform the role of aid in the region. However, if evaluations and their findings are doubted, there is no real, reliable way to determine if aid is effective. Finally, because international development initiatives in Latin America are often criticized as exploitative and extractive, it is important to ensure these evaluations are not overstating or misinterpreting results, and that projects are, at the least, doing no harm (Lang and Mokrani 2013).

Hypotheses & Expected Observations

Hypothesis 1: Evaluations appear to employ technically-sound procedures but, on closer analysis, actually prioritize form over substance. Further, I predict significant variation in structure and quality across reports and across evaluators.

> <u>Expected Observations:</u> On the individual evaluation level, methods are consistent, and structure is appropriate. But, findings do not truly, or only weakly, answer the questions they asked. The central questions of impact and/or performance remain unanswered. Further, when evaluations are cross-compared, significant structural differences will emerge, and some evaluations will emerge as superior to others.

Hypothesis 2: Evaluations sponsored by the MCC overstate the impact and successes of MCC projects. Evaluators, although supposedly independent, feel beholden to the expectations of the MCC.

> <u>Expected Observations:</u> Secondary reports do not affirm the findings found in the MCC-sponsored reports and/or suggest that the MCC-sponsored reports have overestimated

findings. Evaluators indicate through interview answers that they did not feel truly independent when conducting evaluations.

Hypothesis 3: MCC evaluations are donor-centric rather than beneficiary centric, prioritizing the financial, result-based needs of donors rather project impact on beneficiaries.

Expected Observations: Evaluations are structured and worded to appeal to the economic priorities of priorities of donors, particularly emphasizing macroeconomic growth over the benefits experienced by individual beneficiaries. Effects felt by beneficiaries such as improved quality of life or poverty eradication are neglected in favor of macroeconomic advances felt at a national level.

Methodology

Why the Case Study?

A case study enables me to draw conclusions across the MCC's three Latin American compact projects and their respective evaluations. My case study will include varied methods, including a document analysis of the evaluations from these projects and interviews with the creators of these evaluations. By both analyzing the evaluations themselves and speaking to individuals intimately acquainted with the evaluation process, I will be able to explore these intricacies of these documents and illuminate my findings with human perspectives perhaps not discernable from my preliminarily analysis.

The evaluations I will study represent the entirely of those published for the MCC's Latin American compact projects. Again, these compact projects are 5-year, large-scale grants to countries that satisfy all of the MCC's eligibility criteria. Since its inception, the MCC has completed three compact projects in Latin America – the El Salvador Compact (2006), the Honduras Compact (2005), the Nicaragua Compact (2005).

I have chosen to analyze only the evaluations for the MCC's completed compact projects. I excluded the MCC's completed threshold projects from my analysis because the MCC seldom sponsors substantive evaluations of these short-term and small-scale projects. Finally, I excluded the MCC's ongoing and recently-completed projects in Latin America – the El Salvador Investment Compact and the Guatemala Threshold program – because evaluation efforts are incomplete. I did, however, apply my findings to these projects, making recommendations to improve project trajectory, sustainability, and eventual evaluation efforts.

Document Analysis

To conduct my document analysis, I used compact evaluation documents created by independent proprietors and several other evaluators. I analyzed three types of evaluations: impact evaluations, performance evaluations, and multiple-method evaluations. While the interpretation of findings considers the different goals of each evaluation type, methods remained constant. Table 4 details the 18 evaluations included in the document analysis, as well as evaluation type, year, and corresponding letter.

Each of these 18 evaluation documents corresponds to – and evaluates – a compact sub-activity. The overarching country compacts are comprised of these sub-activities. Specifically, the Honduras compact has three sub-activities (and three evaluation documents), the Nicaragua compact has six sub-activities (and six evaluation documents), and the El Salvador compact has nine sub-activities (and nine evaluation documents). The names of these sub-activities – which are also the names of the evaluation documents – will be henceforth designated in italics.

I examined and assessed evaluation documents according to five qualities, defined below:

1.) <u>Concurrence</u>: Evaluate the concurrence between project goals and findings of the evaluations (Hypothesis 1)
2.) <u>Consistency</u>: Evaluate the consistency of the metrics used throughout the evaluation, as well as these metrics' appropriateness given original project design (Hypothesis 1)
3.) <u>Accuracy</u>: Evaluate the concurrence between findings of the evaluations and the findings of available secondary sources (Hypothesis 2)
4.) <u>Variant Structure</u>: Evaluate the significance of the variation in evaluation authors (and their respective agencies) on evaluation structure and findings (Hypothesis 1)
5.) <u>Inclusive Focus</u>: Evaluate the extent to which the evaluation considers project beneficiaries – and the ultimate goal of the MCC (poverty reduction) – while conducting the evaluation and drawing conclusions. (Hypothesis 3)

To score the evaluations, I used a 1-5 scale to rank the extent to which evaluations satisfied the five abovementioned steps. In this ranking system, 5 = excellent; 4 = good; 3 = fair; 2 = poor; and 1 = bad. I have integrated the average evaluation score according to step of the document analysis in the Results section.

This scoring system is relative, as evaluations are compared to each other, rather than an outside standard. And, although the scoring process is inherently subjective, I used a clear scoring rubric that ensures consistency and enables replication (see Appendix 1).

To actually perform each of the five steps in my analysis, I used separate processes. To execute step one, I manually compared the original project goals listed in the evaluations and the findings of the evaluations. High-scoring evaluations present findings according to these original project goals.

Step two consists of tracking the consistency of metrics used throughout the evaluation process. If evaluations state they will measure the extent to which household incomes improved after project implementation, for example, high-scoring evaluations actually do this – presenting analyses and findings according to the originally-established metrics.

Step three consists of cross-comparing the findings of the MCC-sponsored evaluations to existing secondary sources that also evaluate the El Salvador, Honduras, and Nicaragua compacts. If the findings of secondary evaluations and those of MCC-sponsored evaluations are concurrent, for example, such concurrency strengthens the case for the accuracy of the latter-mentioned evaluations. Available secondary evaluations include annual Congressional Research Service reports and annual Government Accountability Office reports. Unfortunately, these secondary evaluations were relatively scarce, and I was unable to directly compare the findings of each of the MCC-sponsored evaluations to secondary evaluations. As such, step three addresses only the findings of the Honduras *Transportation*, *Farmer Training*, and *Agricultural Public Goods* evaluations (each of the three Honduras compact sub-activities) and the Nicaragua *Transportation* evaluation (one of six total Nicaragua compact sub-activities).

Step four consists of comparing the evaluations, specifically the structure, user-friendliness, and clarity of these. I paid special attention to differences in structure across evaluation proprietor, as well as differences between impact and performance evaluations. This comparison enabled me to determine best practices for evaluations. This step also enabled me to determine how the evaluations process can be improved.

Although I will expand upon the question of the fifth step in my interviews with the evaluators, a manual search for the ways in which evaluations consider project beneficiaries and poverty reduction – or fail to do so – is also instructive. To determine the extent to which beneficiaries were considered in these evaluations, I paid special attention to the involvement of stakeholders and beneficiaries in the evaluation processes as well as the extent to which evaluations measured project impact on beneficiaries.

I organized the findings of the document analysis according to the components of the hypotheses and research questions. To evaluate whether the evaluations were sound in content and structure (Hypothesis One), I examined the Concurrency (Document Analysis Step One), Consistency (Step Two), and Variant Structure (Step Four) of evaluations. To examine whether the MCC-sponsored evaluations overstated findings (Hypothesis Two), I discussed the Accuracy of MCC-sponsored evaluations (Step Three). Finally, to examine whether evaluations are donor-centric or beneficiary-centric (Hypothesis Three), I discussed the Inclusive Focus of evaluations (Step Five), commented on the user-friendliness and language used in the evaluation documents (Step Four), and examined the metrics prioritized in the evaluations (Step Two).

Interviews

To supplement the document analysis, I interviewed the authors of the evaluations. In several cases, evaluators were involved in multiple project evaluations. As such, I conducted only ten interviews but was able to speak with someone involved in each of the 18 evaluations. I conducted and recorded the interviews over Zoom and uploaded the recordings to a transcription software called Otter.ai. I then edited the transcripts for clarity and coded them thematically using an Excel database (henceforth called the Analysis Document).

Interviews included a common set of questions and project-specific probes. Standard questions sought information about the evaluation process – the intricacies of how evaluations are conducted from start to finish. Project-specific probes focused on the difficulties evaluators faced in data collection, the extent to which beneficiaries and stakeholders were included in the evaluations process, and the metrics used in the evaluations. For example, when speaking to evaluators involved in the Nicaragua projects, I asked how their evaluations were affected by the MCC's termination of the Nicaragua compact. I also inquired about the perceived limitations and impacts of the evaluations process and how it can be improved in the future (see Appendix 2).

I used insights from the interviews to elaborate on the technical and structural soundness, accuracy, and donor-centricity of the evaluations. This interview data not only complements the "story" of the document analysis results but also illuminates that which is not apparent in the evaluation documents. I implemented the interview data according to relevant hypotheses in both coded and interpretive manners and particularly emphasize the evaluations process and its impacts, the inclusion of beneficiaries and stakeholders, and the potential for learning from the evaluations.

Empirical Findings

The results of the document analyses both confirmed and disproved components of the hypotheses. Despite variation among evaluations, the quality of these evaluations was better than anticipated. The average evaluation was technically sound, and secondary evaluations did not directly contradict the results of MCC-sponsored evaluations (Hypotheses One & Two). However, the results did confirm that the evaluations evidence some donor-centricity and can improve in their inclusion of beneficiaries (Hypothesis Three).

The average evaluation is sound in technical procedures and in structure

Of the 18 evaluations included in the case study, the average evaluation is sound in technical procedures and structure (Hypothesis One). The examination of the concurrency between original project goals and evaluation findings indicates that the average evaluation evidenced good to fair concurrency. The examination of the consistency of metrics used throughout the evaluations indicates that metrics were mostly consistent. Finally, the assessment of evaluation structure and clarity indicates variation according to evaluation agency and proprietor. However, this assessment also indicates that the average evaluation was clear and has a well-devised structure. Evidence from interviews also supported the technical and structural soundness of the evaluations.

Most evaluations evidenced good concurrency between original project goals and evaluation findings. The average score was 3.66, a fair to good score. However, the evaluations' levels of concurrency varied (Figure 5). Five evaluations received the highest score of five, four evaluations received a four, seven evaluations received a three, two evaluations received a two, and no evaluations received the lowest score of one. As the majority of evaluations received either an excellent, good, or fair score, this suggests that original project goals and evaluation findings were concurrent in the evaluations.

The characteristics supporting concurrency between project goals and evaluation findings were a clear and early description of project goals and a presentation of findings according to project goals. For example, the original goals of the El Salvador *Water and Sanitation Project* were improved access to potable water, increases to household income, reduction of time spent seeking water, and reduction of morbidity from water-related diseases. The evaluation presented findings according to these exact project goals, even when findings were null (Doc. J 2014).

The characteristics of evaluations that did not exhibit great concurrency were a vague or late description of project goals and a presentation of findings incongruent with project goals. For example, the Honduras *Transportation & Farm-to-Market Roads* Project evaluation presented its goals of reduced transportation costs, upgraded roads, reduced price of consumables, and increased access to social services in a clear manner. However, its presentation of findings did not elaborate on these goals and instead asserted broad statements such as, “the Transportation Project shows some statistically significant effects on a variety of indicators (income and travel times to places of interest), those impacts are very small” (Doc. B 2013, pg. 71).

Evaluations also evidenced good consistency in metrics, with an average score of 3.94. Despite this high average, levels of consistency varied across the 18 evaluations (see Figure 6). Six evaluations scored a five, six evaluations scored a four, five evaluations scored a three, and one evaluation scored a two. No evaluations received the lowest score of one. Again, the majority of evaluations received either excellent, good, or fair scores, which is a testament to the consistency of metrics in these evaluations.

Evaluations that exhibited good consistency described metrics early in the executive summary, aligned metrics with original project goals, and presented findings according to original project goals and metrics. In these evaluations, the body of the text was also organized according to original metrics. The Nicaragua *Transportation Project* evaluation is a good example of these characteristics. Its metrics – the economic rates of return (ERR) for the updated roads, changes in household consumption, and changes in the availability and cost of goods –

were listed and explained early in the evaluation. The body and calculations of the evaluation were organized according to these metrics, and the findings were presented according to them as well. Finally, these metrics are appropriate given the goals of the project, which were the reduction of transportation costs and increased household access to goods (Doc. I 2014).

Evaluations that did not exhibit good consistency presented methods vaguely or did not remain consistent with the use of metrics throughout the evaluation body and conclusion. Further, in evaluations with inconsistent metrics, the correlation between project goals and metrics was not always apparent. For example, the Honduras *Agricultural Public Good Grant Facility* evaluation did not describe metrics early in the evaluation text, nor organize calculations according to these metrics. There is no true summary of findings. Further, because the evaluation does not describe original project goals, it is difficult to determine the appropriateness of the metrics used in the evaluation (Doc. C 2010).

Finally, the average evaluation had good structure and clarity, with an average score of 3.55. However, no two evaluations had identical structures, as structure varied by evaluator and agency. The evaluation scores reflect this variation (see Figure 7). In this case, three evaluations received a five, nine evaluations received a four, two evaluations received a three, three evaluations received a two, and one evaluation received the lowest score of one. Although the average evaluation had good structure and clarity, the variation in scores indicates that the evaluations – when cross-compared – differed in quality.

Such differences in quality were even apparent at the country and evaluator levels. For example, some El Salvador evaluations were superior in structure than others (see analysis for document J versus that of documents L and Q, included in Analysis Document). The same structural differences were true for certain Nicaragua evaluations (see analysis for document D versus that of documents E and F). Further, the same author or set of authors produced evaluations of different qualities, another testament to the variation within the case study (see differences between documents K, L, and M). In fact, certain evaluation agencies consistently produced better-structured evaluations than did others. For example, the evaluations conducted by a first agency scored the collective highest average, with an average structure score of 3.75 across its four evaluations. A second agency followed, with a 3.2 average structure score across its five evaluations. Last in structure was a final agency, with an average score of 3 across its three evaluations. Interestingly, the best-structured evaluations – those conducted by the first agency – were the most recent, and the worst-structed evaluations – those conducted by the final agency – were the oldest. The middling evaluations of the second agency were conducted in the years between the first and final evaluations. This apparent relation between recency and superior evaluation structure may suggest that evaluation structure improved over time.

The most clear and well-structured evaluations followed the organization of an academic paper and attempted to simplify dense methodologies and calculations. These evaluations contained clear executive summaries, introductions, methodologies, data analyses, and conclusions. Some evaluations even contained literature reviews, which helped to situate evaluations according to similar studies and motivate the importance of the findings. The structure and clarity of the El Salvador *Water & Sanitation Project* evaluation was particularly effective, including a literature review and parallel sectioning and wording throughout (Doc. J 2014). This evaluation even used plain language to describe complicated methodologies and analyses, a testament to its clarity.

The least-clear and most poorly structured evaluations contained vague executive summaries, were organized unintuitively, or were incumbered by technical jargon. For example, the organization of the El Salvador *Rural Electrification Project* was difficult to follow, as it did not contain an introduction nor sufficiently describe the project and its goals. The body of the evaluation was difficult to follow and was encumbered by summary and thematic tangents. Furthermore, this evaluation embedded complex equations without providing sufficient explanatory text. The clarity and comprehensibility of the evaluation suffered as a result.

For the most part, evidence from interviews with evaluators confirmed that evaluations are technically sound. To ensure the quality of evaluation design and findings, project evaluators work with MCC M&E personnel. Interviewees described this review process as particularly thorough, stating that the MCC was concerned with the rigor of findings and the accuracy of assumptions. Evaluator B's experience with the MCC review process speaks to this rigor:

> "Hubo un dialogo cercano [entre nosotros y el MCC], intercambiando y afinando detalles [sobre el diseño de la evaluación]…Nosotros encontramos [que el proyecto tuvo] impactos fuertes, y MCC quería asegurarse de que estos fueron resultados verdaderos…Nos hicieron pasar incluso por mas rondas de revisión de los que habían sido planeadas inicialmente."
>
> There was a close dialogue [between we evaluators and the MCC], interchanging and honing details [of the evaluation design]…We found that [the project had] strong impacts, and MCC wanted to make sure that these results were true…They made us perform even more rounds of revision than what had initially been planned.

Evaluator B also noted that the MCC hired experts in the project subject matter to give findings a secondary review. Another evaluator, Evaluator A, also mentioned having participated in one of these secondary reviews, a testament to the thoroughness of the MCC's review process.

Other evaluators, however, described the MCC review process as less thorough. Evaluator F, for example, suggested that the MCC's level of concern with evaluation design depends upon the size and importance of the project being evaluated. This explanation seems plausible, as several of the evaluators who did not endorse the thoroughness of the MCC review process were those assessing small programs through performance evaluations. Furthermore, these evaluators only expressed concern about MCC thoroughness in the design phases of projects and felt that the MCC sufficiently reviewed final results. Although these accounts call into the question the rigor of the MCC review process, the majority of evaluators participating in impact evaluations suggest that the review process and technical components of the evaluation are sound and thorough.

Interview data also supports the variation in evaluation structure found in the document analysis. Notably, only one evaluator, Evaluator F, mentioned receiving a template from the MCC regarding evaluation structure. This discrepancy with the other evaluators could be explained by the fact that Evaluator F conducted the most recent evaluation. As such, the MCC could have developed evaluation guidelines and templating in the years between the early projects and the most recent. Evaluator H affirms this suggestion, describing the early years of evaluation as an experimental process: "We started working with [the MCC] when they were a young agency. In many ways, they were developing processes as we went…our evaluations were a testing ground." Such a dynamic evaluation environment could help explain the differences in structure observed in the document analysis.

Another possible explanation for the variation in evaluation structure was that the MCC was very permissive of diverse evaluation designs. In one case, Evaluator J described pioneering a performance evaluation design. The MCC swiftly accepted this new design with little revision, and three of the evaluations in the case study are conducted according to this design (Docs. E, F, G 2012). Both the lack of evaluation templating and the tolerance of diverse evaluation design can help explain observed variations in evaluation structure.

Secondary evaluations do not directly contradict the findings of MCC-sponsored evaluations

Available secondary evaluations did not directly contradict the findings of the average MCC-sponsored evaluation. Rather, the secondary evaluations from the Congressional Research Service (CRS) and GAO either prioritized different metrics or presented findings with slightly different figures than did the MCC-sponsored evaluations. Furthermore, interview data suggests that evaluators felt independent while conducting their evaluations and did not feel pressured to publish positive results. This finding goes to support the integrity of the evaluation findings, despite small inconsistencies with those of secondary evaluations. These findings disprove the initial hypothesis, which supposed that MCC-sponsored results were inflated and that evaluators were beholden to the MCC's expectations.

The comparison of MCC-sponsored evaluations and secondary evaluations revealed certain inconsistencies between the two. Given these inconsistencies, the average Accuracy score (concurrence between findings of MCC evaluations and secondary evaluations) was a 3, a fair score. One evaluation received a score of four, two received a score of three, and one received a score of two (Figure 8).[4]

The inconsistencies between the MCC-sponsored evaluations and the secondary evaluations were small numerical differences. For example, while the MCC-sponsored

[4] Note: This assessment of the accuracy of the MCC-sponsored findings does not include all 18 evaluations in the case study, as secondary evaluations were relatively scarce.

evaluation for the Honduras *Farmer Training Project* found "more than 7,500 farmers" to have received training, a CRS evaluation reported this figure as "more than 7,000 farmers"[5] (Doc. A 2013, pg. 8; CRS 2014, pg. 32). The results of the Honduras *Transportation Project* and those of complementary CRS reports also differ. While the MCC-sponsored evaluation asserts that the project updated "65 km of secondary roads and 495 km of rural roads," the secondary evaluation lists these figures as "45 km of…secondary roads…and 422 km of rural roads" (Doc. B 2013, pg. 2; Tarnoff 2014, pg.32). However, these discrepancies do not necessarily suggest that the MCC-sponsored evaluations flagrantly inflate results.

Evidence from interviews supports the academic integrity and independence of evaluators. Each of the evaluators said that they felt independent in the evaluations process. They felt that their data and findings were their own and did not feel pressured by the MCC to publish positive results. This perspective is reinforced by the evaluations themselves, which often assessed projects as having very small (Docs. B 2013; D 2012; E 2012; I 2014) or even negative impacts (Doc. R 2014). The MCC's rigorous review process described in the previous sub-section also supports the integrity and accuracy of findings, especially for large-scale evaluations that measured project impact with counterfactuals. These findings contradict the original hypothesis, which supposed that evaluators would feel beholden to the expectations of the MCC and would not be sufficiently independent. The reported independence of evaluators instead goes to support the accuracy and integrity of MCC-sponsored evaluations, despite their occasional inconsistencies with secondary reports.

[5] Note: Technically, these figures could be consistent.

Evaluations can improve in inclusion of beneficiaries

Of the evaluations included in the case study, the average evaluation evidenced donor-centricity – prioritizing macroeconomic trends and accountability measures (such as returns to investment) rather than individual-level impact. The analysis examining the extent to which evaluators included and prioritized beneficiaries suggests an exclusion of beneficiaries. This analysis also suggests that these documents fail to prioritize learning. Next, the examination of the user-friendliness and language of the evaluations indicates that evaluations employ elite language and are inaccessible to the layperson. Finally, the assessment of the metrics used in the average evaluation indicates a prioritization of donor-centric rather than beneficiary-centric metrics. Evidence from interviews with the authors of the evaluations supports these findings. I will expand upon these findings in the following paragraphs.

Evaluations evidenced poor inclusion of stakeholders and beneficiaries. The average score was 2.94, the lowest average of all scoring categories. No evaluations received the highest score of five, five evaluations received a four, eight evaluations received a three, four evaluations received a two, and one evaluation received the lowest score of one (see Figure 9). The majority of evaluations received fair, poor, and bad scores, which evidences poor inclusion of beneficiaries.

Despite the low average observed across the 18 evaluations, five evaluations still received a score of four. These evaluations actively involved beneficiaries in the evaluations process and prioritized project impact on beneficiaries in the evaluation texts. For example, certain evaluations featured interviews with both stakeholders and beneficiaries, while others included focus groups. One of these is the El Salvador *Formal Technical Education Project* evaluation, which includes detailed interviews with the students and administrators involved in the education programming. Through these interviews, beneficiaries were able to both voice their perceptions of the project and provide recommendations for future projects. The evaluation text also prioritizes project impact on beneficiaries by integrating individual-level metrics such as student graduation rates, school attendance, and academic achievement with macro-economic metrics such as labor market outcomes.

The least-inclusive evaluations were those that make little effort to involve beneficiaries in the evaluations process nor incorporate individual-level wellbeing in the evaluation texts. Such evaluations typically leveraged household surveys, which prioritize consumption and income levels instead of beneficiary feedback. The El Salvador *Connectivity Project* is one such evaluation and uses only household surveys to determine project impact on beneficiaries. The body of the evaluation reflects these priorities, using only metrics of consumption and income to speak to beneficiary wellbeing and project impact. Although the measure of – and positive changes in – income and consumption can have significant implications for those living in material need, the sole reliance on these indicators precludes beneficiary participation.

Furthermore, evaluations evidenced a minimal prioritization of learning. In many cases, the "Learning" or "Recommendation" sections in these evaluations are no longer than a page,[6] and many of the recommendations focus on improved data collection strategies rather than ways in which project implementation might improve. The official MCC responses to these learning sections are also brief – they are seldom more than three pages of project summary followed by a brief, bullet-pointed list of recommendations. Moreover, these documents make no promises that recommendations will be considered or implemented.

[6] Note: Other sections are typically longer and more substantive.

Evaluations evidenced poor accessibility, as they employed elite language (such as econometric terminologies), contained dense calculations, and were rarely available in Spanish. In fact, only one of the 18 evaluations, that of the Honduran *Agricultural Public Goods Project*, was translated. If in-country personnel, stakeholders, or beneficiaries cannot read the evaluations, their capacities to learn and grow from these documents are diminished. Although several evaluations made efforts to simplify methodologies and calculations, the intricacies of econometric analysis still require an elite lexicon and comprehension of academic terminology. The monolinguistic and complex character of these evaluations evidence their inaccessibility to a layperson.

Finally, evaluations evidenced a prioritization of macroeconomic metrics (such as aggregate estimates of increases in income and consumption or market-level indicators), rather than individual or village-level metrics. Certain evaluations do assess individual-level metrics such as time allocation and access to services such as clean water, education, healthcare, and commercial markets. However, these metrics were seldom the priority of the evaluations. Instead, macroeconomic metrics such as economic rates of return, loan repayment rates, and agricultural production levels were the true priorities. This is especially true for the impact evaluations of the case study, which spend millions to establish counterfactuals and calculate returns to investment. Although macroeconomic metrics are central to ensuring donor accountability and the responsible disbursement of funds, their sole use struggles to portray wellbeing on an individual level. As such, an integration of macroeconomic metrics with individual-level metrics is needed to depict a fuller picture of project impact on beneficiaries.[7]

[7] Note: I will expand upon this comment in the subsequent Discussion section.

Interviews evidenced that the evaluations process is exclusive to beneficiaries and prioritizes macroeconomic results. When asked whether beneficiaries or stakeholders were involved in the evaluations process, only four of ten evaluators described a meaningful inclusion of these groups. Of the six evaluators who did not describe meaningful inclusion, two (Evaluators B and F) described that there was “virtually no involvement of stakeholders.” Of the other four evaluators, two (Evaluators C and E) described the involvement of beneficiaries through household surveys, and the remaining two (Evaluators A and D) suggested that implementation staff had involved beneficiaries before project implementation.

For the most part, the exclusion of beneficiaries was not deliberate and instead occurred as a result of financial and logistical constraints. In fact, three evaluators wished they could have better-included beneficiaries in the evaluation process, citing such logistical and financial constraints for their inabilities to do so. While Evaluator E said, “there was a very strict budget for data collection, and I maxed it out with [household] surveys. That being said, it would have been useful to do more qualitative work,” Evaluator J noted that, “the field work was limited, and if they’d given me three or four more days, I would have been able to go out to farms…or visit some people [who benefitted from the projects].” These testimonies suggest that beneficiary groups can be meaningfully included in the evaluations process given a financial and logistical prioritization of these groups.

Another difficulty in including beneficiaries in the evaluations process was a cultural and language barrier. Fortunately, nine of the ten evaluators spoke Spanish and were able to communicate with in-country staff and other stakeholders. Seven of ten evaluators also had experience in the region, and two of the seven are Latin American nationals. The remaining five gained experience in the region through Peace Corps, academic research, or through evaluations of similar development projects. Two of the remaining three had a cultural understanding of the region, while the remaining evaluator had no experience in Latin America. Despite evaluator experience in the region, Evaluator F still emphasized the difficulties associated with engaging beneficiaries, saying “I think [interaction with beneficiaries] would have been smoother if one of us were Salvadoran.” Evaluator I also emphasized the general importance of cultural competency in the evaluations process:

> “[After] having five or six projects in [one country], I [finally] understood how the country worked. Once I started feeling like I knew people and connected with people, I had a much better chance of doing something that's useful and successful. In every country you work in, everything must fit within culture and norms.”

Fortunately, the majority of evaluators involved in these evaluations possessed such regional understanding. However, meaningful interaction with beneficiaries still proved challenging in certain circumstances.

Despite the observed constraints and barriers to beneficiary inclusion, evaluators still emphasized the importance of inclusive evaluations processes. Evaluator H particularly did so:

> “If evaluators aren’t talking to direct participants about their experiences [with the projects], they might as well be interfacing with an avatar…It's very important to have key stakeholder interviews and focus groups with the direct participants…[Without these], you're missing a huge slice of the story.”

These sentiments suggest that the evaluations process can benefit significantly from the inclusion of beneficiaries, despite the extra resources required to meaningfully engage these groups.

Interview evidence suggests that the potential for learning from evaluations is minimal. Evaluators described a formal review and presentation process for finished evaluations, which often occurred both at MCC headquarters and in recipient countries. Evaluator H acknowledged this commitment to accountability, saying “The MCC does have its faults, but I appreciate that they actually want to be told the truth about whether [projects] worked.” However, Evaluator E questioned the MCC’s ability to actually apply findings to future projects: “sometimes there's lots of potential for learning and to impact program design, sometimes evaluations just go in the garbage.” Evaluator H was even more critical, saying, “in the end, there is no intentional activity that goes into learning or applying the lessons. [The evaluations] just go into the file.” Finally, Evaluator I suggested that the MCC flagrantly ignores learning from evaluations: “[The MCC] just comes in [to countries] and keeps pulling out the same program. [Evaluators] keep saying [the programs] are not working, and [the MCC] doesn’t want to hear it.” Despite a formal review process for evaluations, these testimonies suggest that the MCC capacity to implement learning is variant at best and negligent at worst.

Finally, interviews provided evidence that the evaluations process prioritizes macroeconomic, donor-centric metrics rather than holistic metrics measuring beneficiary wellbeing. Several interviewees emphasized the MCC’s commitment to large-scale impact evaluations that assess project impact through counterfactuals. Evaluator J addresses this attachment to impact evaluations by comparing his work in performance evaluations:

> “The MCC was doing a lot of impact evaluations, which cost millions of dollars, and they saw what I had been doing, which was $20,000 for three evaluations, as opposed to 6 million for one. [My method] was pretty boilerplate in my mind, but it was something [the MCC] hadn’t used because they were doing everything else with economic analysis.”

Evaluator F confirmed this sentiment, noting that the MCC "puts a premium on quantitative data." However, even the authors of evaluations recognized that a reliance on quantitative results and counterfactuals cannot provide all the necessary answers: "some of the reasons why [projects] do work or don't work just aren't part of the data, and you need to talk to people to figure out why," Evaluator J mentioned. Another evaluator, Evaluator H, even noted that "stakeholders were just as happy to hear [performance evaluation] results as they were to hear the randomized control trial results." This quotation suggests that both impact and performance results have value in the eyes of stakeholders, despite their differing uses of quantitative, macroeconomic methods. As such, and despite their demonstrated reliance on indicators such as income and consumptions, the MCC should consider the increased integration of inclusive, beneficiary-focused metrics.

Secondary findings – Project success rate and the use of MCC funding

The document analysis and interview findings demonstrate that the MCC does a quality job in evaluating its projects, despite the fact that evaluations vary in structure. The average evaluation was technically and structurally sound, and secondary evaluations did not directly contradict the findings of MCC-sponsored evaluations. Interviews with evaluators supported these findings. The results of the document analysis and interviews go to answer the first research question – indeed, the MCC has done a quality job in evaluating its Latin American compact projects on the technical and structural fronts. However, the document analysis and interview findings also indicated donor-centricity in the evaluations process and suggested that this process can better-include project beneficiaries.

The actual findings of the evaluation documents have important implications for the MCC's presence in Latin America, as they speak to the success rates of the projects and whether project expenditures were proportional to project impact. While evaluation findings indicate that only half of projects experienced significant success, a cost-impact comparison of the transportation projects (the most expensive), suggest that benefits were not felt at an individual, household level. Evidence from interviews confirms these points. These secondary findings also have implications for the design and evaluation of future development projects in the region (Research Question 2).

MCC Latin American compact project success rate

Evaluation findings demonstrate that only half of the 18 evaluations included in the case study experienced significant success. However, evaluators qualified six of these successes. For example, the Honduras *Agricultural Public Goods Project,* which financed agricultural projects, was found to have effects at a macroeconomic level. However, these effects were not observed "in the economic income of producers [involved in the project];" this project benefitted the market but did not directly impact its intended beneficiaries (Doc. C 2010, pg. 25). Another example of qualified impact was the El Salvador *Investment Services Program*, which provided poor producers and enterprises with loans. The program evaluation found that loans enabled producers to invest in their businesses and generate higher earnings. However, monthly loan repayments greatly reduced these earnings. In fact, 84% of loan recipients reported difficulties in meeting their monthly payments (Doc. L 2015). Finally, the El Salvador *Formal Technical Education* Program, which strengthened educational institutions, brought about increased school enrollment. However, this increased enrollment was not found to have subsequent impacts on student employment or income (Doc. M 2016). These projects, although successful according to certain indicators, often struggled to impact beneficiary wellbeing.

Other evaluators qualified project success by questioning the validity of project design and evaluation findings. The *ProNicaragua* Project, which sought economic growth through foreign direct investment, was found to bring about the investments and create jobs. Although the evaluation could not establish causation between the project and job creation, the corresponding evaluator emphasized that the project's data-reporting system relied on "unverified, self-reported data" (Doc. F 2012, pg.14). Further, the Nicaragua *Property Regularization* project, which sought to strengthen property rights, was found to have restored or granted titles to vulnerable populations. However, the evaluator qualified these results, as the project had "too short [an implementation period] to adequately access the validity of the logic's assumptions" (Doc. G 2012, pg. i) Finally, the El Salvador *Rural Electrification Project*, which sought to extend the electric grid to rural areas, was found to increase household income. However, the corresponding evaluator qualified this success as "noisy and not statistically significant" (Doc. Q 2016, pg. xv).

Evaluations found three projects to have heterogeneous impacts, benefitting wealthier participants rather than less-wealthy individuals. These projects were the Honduras *Farmer Training Program*, the Nicaragua *Rural Business Development Services Program*, and the El Salvador *Production and Business Services Program*, which sought economic development through farmer training and technical assistance programs. In each of these cases, high-resource farmers benefitted the most from programming, while smallholder farmers went largely unaffected (Doc. A 2013; D 2012; K 2013). When prompted about the heterogenous results of one of the projects, the corresponding interviewee explained that implementation personnel even attempted to remove smallholder farmers from the project, as they were unlikely to be successful. This evaluator confirmed this “cherry-picking” phenomenon:

> "I felt that the most vulnerable [beneficiaries] were being left behind. We [implementors] were widening this gap, and we cherry-picked [which participants would] do even better…It seems like these programs really benefited people who were most likely to already be doing fairly well.”

As the above observation suggests, these projects functioned only to widen the gaps between wealthy and less-wealthy participants.

Yet other programs were evaluated as minimally successful, or even as complete failures (the remaining 6 projects). The minimally successful were the Honduras *Transportation Project*, the Nicaragua *Transportation Project*, the El Salvador *Connectivity Project,* and the rice component of the Nicaragua *Rice & Plantain Farmers Program*. The Honduras *Transportation Project*, which sought to repair stretches of primary and rural roads, succeeded in doing so. However, the impact of the roads on the average household was found to be low (Doc. B 2013). The Nicaragua *Transportation Project* also sought to repair roads and improve access to markets. However, the overall effects of the project were categorized as modest (Doc. I 2014). The same was true for the El Salvador *Connectivity Project*, (Doc. P 2016) although I will expand further on these transportation projects in the subsequent section. Finally, for rice component of the Nicaragua *Rice & Plantain Farmers Program*, which financed rice production, it was not possible to “reject the null hypothesis that average impacts of the program on yields and revenues were zero” (Doc. H 2015, pg. 3).

Evaluations found the Nicaragua *Farming & Forestry Project* and the El Salvador *Handicraft Value Chain Project* to be failures. The Nicaragua *Farming & Forestry Project*, which provided subsistence farmers with tree seeds to cultivate, grow into lumber, and ultimately sell for furniture production, was overwhelmed by logistical difficulties and poor planning. As a result, the project did not reach any of its reforestation targets, planted trees had low survival rates, and a significant portion of the project, which was designed to remedy water supply constraints, was completely cancelled (Doc. E 2012). The El Salvador *Handicraft Value Chain Project*, which sought to provide handicraft artisans with business skills, technical training, and financial supports, was found to negatively impact artisan salaried income and household income. These results were so negative that the evaluation suggested that future projects should facilitate artisan transition out of the handicraft value chain (Doc. R 2014).

A possible explanation for the minimal successes – and complete failures – of these projects were that they were poorly conceived. When speaking about the failed Nicaragua *Farming & Forestry Project*, the corresponding evaluator emphasized this very point:

> "From an unplanted seed to a sold piece of furniture in three years...[trees] don't even grow that quickly. And then you work with farmers who had never grown anything [other than] corn and beans before. [Logically], they use their good soil to feed their families (because they're subsistence farmers) and then, with their crappy soil, they'll try this little project with the trees. To me, [the realities of the subsistence farmers] was a huge limiting factor [to the project]. The economists don't look at it like that."

This sentiment speaks to the questionable design of certain MCC compacts – this project was so poorly conceived that, even the best of circumstances, it could not succeed. This evaluator went on to question the impact potential of the *ProNicaragua* Project, which sought to bring about economic development and job creation through foreign direct investments:

> I'm less convinced about the foreign direct investment actually trickling down, especially in a place like Central America. The winner grabs, and the people that get the jobs are given poor salaries and no livable wage or health care.

In addition to poor project conception, this evaluator suggests, some projects might actively lack the capacity to meaningfully affect their intended beneficiaries.

Despite the above-described qualified successes, heterogenous and minimal results, and questionable project design of the MCC's Latin American compact projects, some of these projects brought about important advances in infrastructure, education, electrification, sanitation, and in the provision of skills and services (Doc. J 2014; Doc. N 2015; Doc. O 2017). And, these advances had important impacts on beneficiaries. The larger capacity of these projects to bring about impact proportional to project expenditure, however, remains unexamined.

Was project impact proportional to expenditure?

The minimal results of the most expensive sub-projects, the Honduras *Transportation* Project, the Nicaragua *Transportation* Project, and the El Salvador *Connectivity* Project, suggest that project impact might have been disproportional to project cost. In total, the Honduras, Nicaragua, and El Salvador compacts constituted expenditures of $205,000,000, $113,500,000 and $460,940,000, respectively (MCC, *Where We Work*). The *Transportation* and *Connectivity* projects, which sought to reduce transportation costs and facilitate market access by improving roads, constituted over half of these total compact expenditures, with $120,584,457 for the Honduran *Transportation* project, $57,884,159 for the Nicaraguan, and $270,051,380 for the Salvadoran *Connectivity* Project (MCC, *Where We Work*).

Despite mass expenditures, the *Transportation* and *Connectivity* projects were found to have only minimal effects on the individual household. For example, the evaluation for the

Honduras project "determined that the mean household-level impact of the project, averaged over the nation, [was] low" (Doc. B 2013, pg. 9) For the Nicaragua project, improved road infrastructure led to reductions in price of dairy and egg products, which led to a subsequent increase in consumption of these food items. Despite this consumption increase, the project evaluation found that "aggregate estimates of the impacts of the Transportation Project [were], in general, modest" (Doc. I 2014, pg. 83). What's more, only one section of the improved road, "achieved the benchmark 8% rate of return" for direct user benefits (Doc. I 2014, pg. 83). Finally, evaluations for the El Salvador *Connectivity* Project found the improved road to have "modest impacts, if any, on household welfare" (Doc. P 2016, pg. xiv). Although there was "clear evidence of improvement in market access, increases in labor supply…reduction of leisure, and increases in the use of fertilizers, there were no significant changes in agricultural sales, harvests, income, or expenditure" (Doc. P 2016, pg. 81). These multi-million-dollar projects were impactful in their provision of improved infrastructure. Unfortunately, evaluations indicated that the benefits of this improved infrastructure were minimal for the average household, suggesting that beneficiary impact was not necessarily proportional to expenditure.

The high expenditures and low returns of mass infrastructure projects such as the Honduras, Nicaragua, and El Salvador *Transportation* and *Connectivity* projects beg the question of whether these tens of millions could be put toward a different kind of project – one that is designed to tangibly impact beneficiaries. One evaluator particularly addressed this question of expenditure versus impact while discussing one of the infrastructure projects:

> "So much could have been done with this quantity of money…maybe if a more thoughtful plan had been developed, one that involves stakeholders and looks at the region and the communities that are most impacted by poverty and a lack of social

> mobility…If [the MCC found] ways to target those communities instead of [thinking], 'we're going to raise the country up, and somehow benefit these very poor people,' [so much more could be done]…I think [this mindset] is so flawed, but that is their approach. When you think about the hundreds of millions of dollars these contacts are valued at it's insane. [We can] leverage this much money to do so much more."

The millions expended on these projects, as the above perspective suggests, often fail to bring about proportional impact.

Infrastructure projects such as the Honduras, Nicaragua, and El Salvador *Transportation* and *Connectivity* Projects were conceptualized according to the motto of the MCC – that, through economic growth, poverty can be reduced. As evaluations indicate, however, the benefits of macroeconomic growth seldom trickle down to the impoverished. Fortunately, as the above evaluator suggests, development projects can deepen impact if they are made inclusive and are designed according to community need. These considerations lead to the subsequent section, which includes a discussion of the ways in which the development process in Latin America might improve.

Discussion & Implications

The findings of the document analysis and expert interviews broach five implications – and areas of improvement – for the MCC evaluations process. First, they indicated that the development process could better include beneficiaries. Second, findings emphasized the importance of cultural competence in the development process. Third, results also showed a prioritization of macroeconomic metrics and suggest that such metrics may fail to represent the entire 'picture' of project impact on individuals. Fourth, findings suggested that the MCC could better standardize its evaluations process. Finally, results indicated that the MCC should more actively prioritize and apply learning from its completed projects and evaluation documents.

In the following paragraphs, I discuss these areas of improvement by providing recommendations for the MCC development and evaluations processes. The first three sections – which discuss the importance of beneficiary inclusion, cultural competency, and multi-dimensional evaluation metrics – address the implications of findings for future projects in the region (Research Question 2). I supplement these discussions with recommendations, primarily structural changes by which the MCC's development process can be made more inclusive, beneficiary-centric, and holistic. I support these recommendations with evidence from the expert interviews, as well as with the perspectives of Latin American scholars.

The final two sections include recommendations to standardize the MCC evaluation process and improve its prioritization of learning, which are also significant to future programming in the region (Research Question 2). These recommendations refer back to the structural variation and meager prioritization of learning observed in the results sections. They also relate to the previously described criticisms of evaluators – that the MCC incorporation of learning from evaluations is variant at best and negligent at worst.

Working toward the meaningful inclusion of project beneficiaries

The need for improved beneficiary inclusion – as the document analysis and interview data indicated – is warranted, as it adds intrinsic value to the development process and brings returns in project success and sustainability (Kliksberg 1999).

Evaluator H emphasized the value of improved beneficiary inclusion in the evaluations process, and suggested that this value is not fully achieved by the current model of household-level surveys:

> “[Evaluators] might fall to too much on the side of household surveys, [thinking], ‘Well, if you did the survey with the farmer or the family then you've got [an understanding of the project].’ But, the field [of development] is pretty clear that you don't understand how [beneficiaries] feel about [projects] or the mechanisms (why [a project] worked or didn't work, or how exactly did impact come about) unless you talk with [beneficiaries]."

This perspective speaks to the practical importance of meaningful beneficiary inclusion – that of the improved quality and findings of evaluations. Evaluator J also discussed the practical benefits of beneficiary inclusion, citing increased project impact and sustainability:

> “If the projects are properly conceived, [with] beneficiary input at the beginning, then they can have a benefit for poor people. And I've seen some winners, but most of what I evaluate doesn't work so well. And, when you get down to the real aspects of whether project will sustain themselves, most don’t.”

This sentiment also emphasizes the tendency of development initiatives – and not just those of the MCC – to fail in their provision of meaningful, long-term impact. However, and as this quote suggests, beneficiary inclusion can go toward remedying these shortcomings.

Latin American development scholars also suggest the benefits of beneficiary inclusion. These benefits result from the premise that, because beneficiaries face development problems (such as poverty and a lack of resources) on a daily basis, they are well-equipped to solve these problems. The first of these benefits is that beneficiary communities can provide development personnel with knowledge of project environments and an understanding of local problems

(Irarrazaval 2005; Chac 2002; Kliksberg 1999). Next, beneficiary participation can reveal to development personnel the most effective and expedient development strategies, as well as the preferences of beneficiary groups (Chac 2002; Serrano 1998; Kliksberg 1999; Guillen 2008). Beneficiary inclusion can also support local institutions and build social capital by providing agency and capacitation (Chac 2002; Guillen 2008; Kliksberg 1999). The inclusion of beneficiary groups can also provide legitimacy and credibility to the evaluations process by providing candid feedback and making recommendations for the future (Chac 2002; Kliksberg 1999; Irarrazaval 2005). Finally, beneficiary participation can – as evaluators suggested – improve project success and impact levels (Serrano 1998; Kliksberg 1999) and facilitate project sustainability (Kliksberg 1999). To ensure these benefits, development personnel should involve beneficiaries in each step of the development process – from project design to an ultimate presentation of evaluation findings (Serrano 1998; Kliksberg 1999; Delgado 2007).

In the same way that beneficiary inclusion can greatly benefit the development process, beneficiary exclusion can harm the process, causing an influx of additional costs (Guillen 2008). For example, excluded beneficiaries may not support development initiatives or even come to resent programming (Kliksberg 1999). Such disfavor or resentment can bring about an underuse of project services (Guillen 2008). As beneficiary inclusion is conducive to project sustainability, beneficiary exclusion can also hinder project sustainability. Poor sustainability may impair agency recuperation of project costs (Guillen 2008). Finally, beneficiary exclusion can engender community-level conflict, including inter-community conflicts of interest, the harboring of impossibly high project expectations, and the consideration of elite interests over those of needy groups (Guillen 2008).

Evaluators also recommended strategies to enhance beneficiary and stakeholder inclusion, including the forging of partnerships with local evaluators and data collection agencies. For example, Evaluator E suggested that the MCC could benefit from “more genuine, equal partnerships with researchers in communities and with the [government] institutions in the countries where projects are being done.” Evaluator F echoed this idea, noting that the MCC should “encourage more local evaluation expertise…, [as] building local evaluation capacity is important." The inclusion of local stakeholders in the evaluations process is another strategy conducive to an inclusive development process.

However, evaluators made important rebuttals to the above arguments – addressing the potentially-extractive potential of beneficiary inclusion and praising the MCC model as more inclusive than other development agencies. Regarding the latter point, Evaluator H highlighted the current MCC structure of recipient country ownership of compact projects:

> “I believe that [the MCC] has a lot of the fundamentals of international development right. The biggest of which is recipient country ownership of the development process itself. [The MCC is] saying, you, [the recipient country], develop your five-year plan, we will give you a whole lot of input into it. But essentially, it's [the recipient country’s] – [they] own it. And [the program is] actually over in five years, so you don't create this overall dependency and inertia toward more and more dependence.”

This account directly contradicts the perspective of Evaluator I, who described the MCC as the “worst” in terms of paternalism and arrogance.

Evaluator F raised another rebuttal to inclusive processes, citing the potentially extractive nature of beneficiary inclusion without a subsequent and coordinated presentation of evaluation findings:

> "It feels very extractive to go into communities and interview them and take [the information] and go do your own thing. But at the same time, I don't know whether the communities would have any use for additional briefings, unless [the briefings] were done in a really coordinated and thoughtful way."

This observation first emphasizes the importance of beneficiary inclusion throughout the development process – including the presentation of evaluation findings – and also reemphasizes the resources required to meaningfully include beneficiaries in the entire development process.

The practical advantages of beneficiary inclusion and above rebuttals lead to the following recommendation:

Recommendation 1: The MCC should more-deliberately include project beneficiaries and country stakeholders in each step of the development process, from project planning, to implementation, to evaluation and presentation of results. The MCC should dedicate a specific portion of project funding to ensure this inclusion, and especially that of non-elite beneficiary groups. To facilitate the inclusion of beneficiaries and stakeholders, the MCC could establish partnerships with local researchers, create open communication pathways between development personnel and beneficiaries (such as focus groups or other qualitative forums), consult beneficiary groups in the project planning, implementation and evaluation phases, and coordinate efforts to present project results to beneficiaries and stakeholders. Furthermore, the MCC should make each evaluation available in the language of project recipients, to ensure the accessibility of these documents.

The importance of cultural competency throughout the development process

Although findings showed evaluators to have good cultural competency – as the majority spoke Spanish and had prolonged experience in the region – evaluators suggested that the development process itself is not culturally competent. They attributed this lack of cultural competence to arrogance and to an unwillingness to spend the resources necessary to adapt development programming, implementation, and evaluation to country context. This lack of cultural consideration has been shown to detriment both project success and sustainability.

Evaluator I particularly emphasized these criticisms, noting that “there was certainly a level or arrogance [in the development process] … [and] America has always been like this, [saying] ‘We know what's best, we’re parachuting in to save the day.’ I think MCC is just kind of the worst at that." Evaluator I continued, citing the unwillingness of development agencies and evaluators to incorporate country context in the development process:

> “None of that [country consideration] happens. There's no time for any of that. [Instead] they just hit the ground running, [saying], “Yep, we've got a program for you!” [Development agencies] want to work fast. And that's why, even if there is a short-term gain once the program [ends], it's not sustainable [and] just slips away. It's like [the agencies] were never there. But, this idea of [development personnel] just dropping in and working in different countries simultaneously, and really [thinking that they] have insights into how [they] might adapt a program is just crazy. I don't care how smart [they] are. It just takes a bunch of time, talking and listening, (and more listening than talking) to really learn how [countries] works. I just don't think people are willing to do that.”

This above reinforces a common theme, that financial and time constrains are adversarial to an inclusive, culturally competent development process. This perspective also supports the practical advantages of culturally competent development, including project sustainability.

Similar to increased beneficiary inclusion, an active consideration of culture has been shown to tangibly benefit the development process, making it more successful and sustainable. Although evaluators suggested that cultural considerations are an obstacle to expedient development, the exact opposite is true. In fact, cultural considerations are theorized as an impetus to sustainable development, as they situate development initiatives according to the need, aspirations, and identity of beneficiary groups (Romero 2015). Another benefit of cultural consideration is that it ensures the comprehensive inclusion of all community groups, even the historically marginalized or those lacking social capital, such as ethnic minority groups, women, and youth (Deruyttere 1997).

In the same way that cultural considerations can benefit the development process, a disregard for culture can have important consequences. These consequences include the community-level rejection and subsequent failure of development programming. Even more severe, however, is that non-culturally competent development, in its pursuit of ‘modernization’ and ‘progress,’ can erase local values, knowledge and culture. In order to prevent these losses and to ensure the conferral of meaningful benefits, development programming must adapt to culture, rather than culture adapting to development (Romero 2015).

The evaluator criticisms of the MCC’s model of development, as well as the evidence of the practical benefits of culturally competent development, lead to the following recommendation:

<u>Recommendation 2</u>: The MCC should allocate specific financial, time, and hiring resources to actively include cultural considerations in compact projects. To facilitate this purposeful inclusion of cultural context, the MCC might contract and assign development professionals according to region-specific experience and language, consult country experts (preferably country nationals) on project design and adapt programming accordingly, and actively consider population sub-sections including indigenous peoples, Afro-Latinos, women, and youth.

The incorporation of multi-dimensional metrics in the evaluations process

Although metrics such as income and consumption have important implications for project impact on beneficiaries – and especially for those living in extreme poverty – evaluators suggested that such indicators may fail to fully depict project impact. Furthermore, the sole use of such metrics effectively equates beneficiary wellbeing with material goods, which has been shown to be a limiting and shallow equation (Pena-Trapero 2009). As such, both evaluators and

Latin American scholars support an integration of multi-dimensional metrics with material metrics such as consumption and income (Romero 2015; Palomino and Pardo 1999). This integration can engender evaluations that more thoroughly and holistically depict project impact on beneficiaries.

Evaluators confirmed both the limitations of income-based metrics and the potential for multi-dimensional metrics, including the access to education, healthcare, increased leisure time, and the like. For example, Evaluator C emphasized the importance of non-consumption metrics for beneficiary wellbeing: “it's a good starting point to look at consumption measures, but there can be other metrics including access to health care [and] education... that may not rely so much on income changes. Certainly, we all believe education can improve wellbeing." Evaluator G spoke to the practical benefits of multi-dimensional, integrated metrics: "[Integrating] qualitative [metrics] throughout [the evaluation] allows researchers to better design instruments, and quantitative metrics can add some context to these. This gives the evaluations balance." Evaluator A echoed the potential for integration, saying that “[he’d] like to think [he’s] not stupid enough to believe that material metrics are the only thing” but later acknowledged that, “a lot of these [consumption-related and holistic] dimensions ultimately interact, and you're probably being foolish if you don't think the economics matter. [Ultimately], it's really thinking about how [both kinds of metrics] interact.” These quotes make a strong case for both the integration of more-holistic metrics and the capacity of these to add value to evaluations.

The integration of multi-dimensional metrics in the evaluations process has a practical use – indeed, not all components of a quality life can be measured according to material accumulation (Pena-Trapero 2009; Palomino and Pardo 1999). In other words, an individual living in material abundance may perceive his life to be of low quality, as it is devoid of social networks or culture (Palomino and Pardo 1999). The converse may also be true – an individual living in a state of material deprivation may perceive her quality of life to be quite high. Such suppositions contrast Western, capitalistic modes of accumulation. However, and especially for indigenous groups in Latin America, accumulation does not define wellbeing (Deruyttere 1997). The indigenous philosophy of *buen vivir* particularly emphasizes this point, as it values a state of wellbeing that extends beyond the material, and particularly emphasizes spirituality, equity, and identity (Villalba-Eguiluz 2017). Unfortunately, none of the MCC evaluations make active considerations of Honduran, Salvadoran, or Nicaraguan indigenous groups, although these groups were likely both direct and indirect beneficiaries of MCC programming[8].

In order to reflect an inclusive and holistic depiction of wellbeing, MCC evaluations should incorporate multi-dimensional metrics. These metrics should be inclusive to diverse cultures and lived experiences and should endeavor to encompass differing definitions of wellbeing (Razo-González et. al 2014; León 2011; Guillen 2008; Villatoro 2012; Actis di Pasquale 2008). Inclusive, multi-dimensional evaluations should examine both tangible and intangible metrics, including personal relations, social networks, and community context and cooperation (Razo-González et. al 2014; Romero 2015). An examination of wellbeing according to both multi-dimensional and traditional metrics can paint a more complete picture of project impact on wellbeing and quality of life (Razo-González et. al 2014).

The demonstrated potential and importance of multi-dimensional metrics in the evaluations process leads to the following recommendation:

[8] Note: Indigenous population constitute 15% of the Honduran population, 5% of the Nicaraguan population, and 7% of the Salvadoran population. More than 90% of indigenous groups are subsistence farmers, a major focus group of MCC programming. Finally, there is a high correlation between indigeneity and extreme poverty (Deruyttere 1997). Given the indigenous population percentages and convergence of MCC target populations (subsistence farmers and high poverty) within these groups, it is reasonable to assume that many program beneficiaries were indigenous.

Recommendation 3: The MCC should prioritize the use of multi-dimensional metrics in its project evaluations. To facilitate the inclusion of these metrics, the MCC could require evaluators to incorporate multi-dimensional metrics with material metrics, encourage evaluators to consider differing definitions of wellbeing and quality of life, or encourage the use of performance evaluations, which already rely more heavily on multi-dimensional metrics.

Standardizing evaluation structure and quality

As briefly mentioned in the results section, the observed variation in evaluations is likely explained by the fact that the MCC contracted dozens of individuals from differing agencies and academic backgrounds to conduct evaluations. Certainly, a diversity of evaluator perspective is not a detriment to the evaluations process, but rather infuses it with nuance. Despite the variance in the evaluators themselves, however, the MCC can still standardize the products of the evaluators – the evaluation documents. Such a standardization in evaluation form may also serve to standardize evaluation quality. For example, the logic of evaluations, the consistency of metrics, and the presentation of findings was more easily discernable in well-organized, clear evaluations. However, the components and findings of evaluations were often muddled by poor organization or cumbersome wording, which ultimately worsened evaluation quality. The observed variation in evaluations leads to the following recommendation:

Recommendation 4: The MCC should endeavor to standardize its project evaluations. To do so, the MCC might standardize evaluation documents by developing a template for these. This template would not specify evaluation content (and would consequently leave room for adaptations to regional context) but would instead create guidelines surrounding evaluation organization and ordering, the language used in introductory and conclusion sections, the visual appearance of evaluations, and the incorporation of graphs and tables in the evaluation document. The MCC could develop this template according to particularly clear and persuasive evaluation documents from former projects.

Seven evaluators mentioned the importance of collaboration between implementor and evaluator personnel for the success of the evaluations process. This observation operates under the logic that, if evaluators are not involved in project design, implementors might design and execute projects that are difficult to rigorously evaluate. Such was the case for the Honduras *Farmer Training* Program, in which evaluators struggled to implement a randomized control trail (because of incompatible project design). Because of these difficulties, evaluators were forced to adapt their methods to a causal modeling approach. The MCC itself has also recognized the importance of collaboration between these two groups – its learning documents suggest that evaluators and implementors should work in "lockstep." These observations of the importance of collaboration lead to the following recommendation:

<u>Recommendation 5:</u> The MCC should remedy the disjoint between implementor groups and evaluators and ensure continued collaboration between these groups in the development process. To facilitate this collaboration, the MCC could actively involve evaluators in the project design and implementation phases by opening communication pathways between the two groups, obliging implementors to consult evaluators on project design, and increasing the frequency and duration of evaluator field visits.

Evaluation documents suggested that difficulties in measuring results emerged if evaluations occurred too quickly after project completion. This was particularly true for the Honduras *Transportation* Project evaluation, which collected data before or immediately after the updated roads were even completed (Doc. B 2013). Such a rushed evaluation timeline may help explain the modest impact of the project. Conversely, the evaluation results were also affected if evaluation processes occurred too long after official programming had ended. This was the case for the El Salvador *Rural Electrification – Solar Panels* evaluation, which found the variety of solar panels used in the project to be technologically obsolete by the time the evaluation was conducted. This evaluation also found many panels to be in disrepair by the time evaluation data was gathered. However, such findings do speak to that project's lack of sustainability (Doc. O 2017). The observed time sensitivity of the evaluations process leads to the following recommendation:

<u>Recommendation 6:</u> The MCC should endeavor to find a 'happy medium' in its evaluations timeline beginning ex-post evaluation only once beneficiaries have had a chance to take advantage of new services, but before the technologies or benefits of projects become obsolete. Such a happy medium can allow evaluators to more-realistically measure project impact on beneficiaries. To ensure the appropriate timeline of evaluations, the MCC could standardize and enforce an evaluation period for each of its projects. To develop an appropriate and effective timeline, the MCC could examine former evaluations timelines to determine which were the most conducive to measuring impact.

Incorporating and prioritizing learning from evaluations

As described in the results section, both MCC-derived learning documents and criticisms from evaluators suggest that the MCC should better-prioritize learning from evaluation documents. Although it seems intuitive to suggest that future development processes should be informed by the successes and failures of prior programming and evaluation findings, both evaluators and the mediocre success rates of projects implemented years apart suggest that this is not the case for the MCC. Rather, as evaluators suggested, evaluations may go "in the garbage." Even worse, one evaluator suggested that failed projects may be implemented repeatedly, despite negative evaluation findings and pleas from evaluators. The need for improved learning processes from evaluations documents leads to the following recommendation:

<u>Recommendation 7:</u> The MCC should better-prioritize learning from its project evaluations, rather than using them only for accountability purposes. To do so, the MCC should continue its formal review of evaluation findings. Next, the MCC should apply these findings, obliging implementors and evaluators alike to read evaluation documents and assimilate evaluation findings – both successes and failures – in future program design, implementation, and evaluation. Projects found to be detrimental to beneficiaries or poorly conceived should either be sidelined completely or widely adapted, rather than being re-attempted in other countries.

Limitations

Several factors limit the present analysis, including the subjective nature of the document analysis, the small sample size of interviewed evaluators, and the fact that some evaluations occurred up to ten years ago, which may have hindered evaluators' recollection of events.

Finally, the analysis is limited because it is based upon evaluation findings, which were conducted and written by fallible human beings working in data-scarce environments.

The analysis is first limited because it was impossible to eliminate all bias in the document analysis. Although the documents themselves were assessed and scored according to a detailed rubric, bias may still exist.

This analysis is also limited by the characteristics of the expert interviews, including a small sample size and evaluators' difficulties in recollecting details from the evaluations process. The expert interviews only included ten evaluators, although more than twenty evaluators were listed on the evaluation documents. This small sample size partially occurred because response rates from evaluators were low. The small sample size is also explained by the fact that, in several circumstances, one evaluator conducted multiple evaluations. As such, with only ten interviews, it was possible to speak with an involved party for each of the 18 projects. The fact that evaluations occurred up to ten years ago also limits the expert interviews. In two cases particularly, evaluators struggled to recall details of the project and evaluations processes.

Finally, the present analysis is limited because it is based upon the data and findings of fallible evaluators. Indeed, evaluation findings were reviewed by the evaluators themselves, by MCC M&E personnel, and, in several cases, by an outside expert. Comparisons with secondary evaluations also enabled the partial verification of MCC-sponsored findings. However, these secondary evaluations were relatively scarce. As such, and although the evaluations process is presented as scientific and objective in nature, it is impossible to fully verify this objectivity, nor the veracity of data. The entire 'truth' of project impact can never be completely identified nor measured, even by the most comprehensive and expensive of impact evaluations.

Conclusion

Since the end of WWII, international development initiatives have been deployed to remedy intractable global problems. Unfortunately, these problems have persisted, despite the billions expended to bring about their remedy. In Latin America specifically, problems of mass inequality, political instability, and exclusion of marginalized groups still persist (UNDP 2019; Banco Mundial 2018; Deruyttere 1997). The United States, with its distribution of neoliberal policies, support of authoritarian regimes, and often ill-advised development initiatives, has been criticized for contributing to such problems (Harvey 2007; Mowforth 2014; Lang and Mokrani 2013). At the very least, the United States has done little to remedy such problems, despite its historic presence and billions expended in the region (Meyer and Gracia 2019; Mowforth 2014).

After decades of development initiatives, the MCC was created to in 2004 to 'do development differently,' focusing on project results and the commitment of recipient governments to their citizens. Since its inception, the MCC has implemented ten projects in Latin America, constituting total expenditures of nearly $1 billion (Brown 2019). The present analysis, through its assessment of MCC programming and evaluations, is significant because it speaks to the successes – or lack thereof – of such programming, as well as the extent to which this programming is being thoroughly evaluated.

Through analysis of the evaluation documents and expert interviews with the authors of the evaluations, the average MCC evaluation was found to be technically sound, although variant in structure (according to evaluation agency). The analysis also shows that secondary evaluations from the GAO and CRS did not directly contradict MCC-sponsored findings, although the former were often more thorough than the latter. Finally, the analysis suggests that evaluations and evaluators could improve in their inclusion of project beneficiaries.

These findings raised important implications for future programming in the region, including the need for and benefits associated with cultural competence and beneficiary inclusion in the development process, as well as the integration of multi-dimensional metrics in the evaluations process. Next, findings suggested the need for improved technical evaluations practices, including the structural standardization of evaluation documents, the collaboration between implementors and evaluators, and the establishment of a 'happy medium' evaluation timeframe. Finally, findings suggested the need for an improved learning process – that evaluation findings should inform future project design, implementation, and evaluation, rather than going "in the garbage." Indeed, much can be learned from the evaluations of the MCC's Latin American compact projects, and this learning should be applied to ongoing and future projects.

As of January 2020, the MCC was implementing two Latin American projects – the El Salvador Investment Compact and the Guatemala Threshold Program. In September 2020, the El Salvador Investment Compact closed. A preliminary press release in September 2020 stated that compact programming was heavily impacted by the COVID-19 pandemic, although the Salvadoran government has pledged to "continue the funding and implementation of the compact's projects" (MCC, El Salvador Press Release 2020). The Guatemala Threshold Program, although still being implemented, will terminate in May 2021. Unfortunately, little time remains for this project. What does remain – for both projects – is the evaluation phase. My analysis and recommendations can have important implications for these evaluations.

After the evaluation phases of the El Salvador Investment Compact and the Guatemala Threshold Program come to an end, the future of the MCC in Latin America in uncertain. Regardless of this nebulous future, the implications of its existing programming are clear. To facilitate a successful development process, the MCC should meaningfully include beneficiaries, actively consider cultural context, widen the metrics used in evaluations, and apply learning to future programming. Until then, the MCC's development and evaluation process will, despite its technical soundness and provision of donor-centric, macroeconomic metrics, continue to fail where it matters most – in the provision of tangible benefits for the individual in need.

Tables & Figures

Table 1 – Baseline indicators used by MCC to determine if countries are eligible for funding

Indicator	Specifications
Be among the less developed countries of the world as classified by the World Bank in its classifications for that year, defined as either	- Fitting the low-income or lower-middle-income category that year (for FY2019, this limit is a Gross National Income [GNI] per capita of $3,895) - Be among the 75 lowest per capita income countries as classified by the World Bank
Not be subject to legal prohibitions on foreign assistance.	- Legislation specifically prohibits aid to the governments of Cuba, Sudan, Syria, and North Korea

(Brown 2019)

Table 2 – Criteria used by the MCC for its Conditional Aid scheme, used to evaluate whether countries should receive funding

Criterion	Definition
Ruling Justly	Promoting democratic governance, fighting corruption, maintaining transparency, protecting property rights, respecting human rights, and creating an enabling environment for civil society.
Investing in People	Providing quality health care and primary education, reducing child mortality, promoting the protection of biodiversity, and taking other opportunities to promote an educated and healthy population.
Fostering Economic Freedom	Fostering enterprise and entrepreneurship, promoting engagement in global trade, strengthening market forces, and respecting worker rights.

(Brown 2019)

Table 3 – Central components of quality evaluations

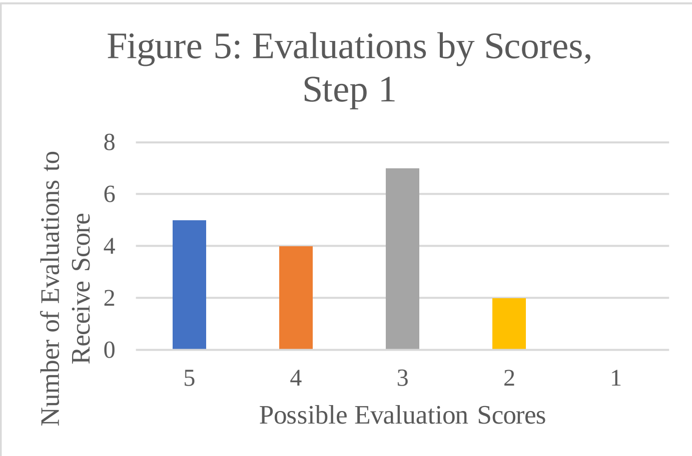
Criterion	Definition
Well-defined scope	The objectives of the information, the purposes to be served and the range of coverage should be clearly set out.
Accurate Data	The data collected should be valid and reliable.
Sound analysis	The analysis of the data collected should be based on robust methodology.
Objective findings, conclusions	The findings and conclusions presented should be supported by the evidence gathered (data and analysis) and should be presented in an impartial (objective) manner

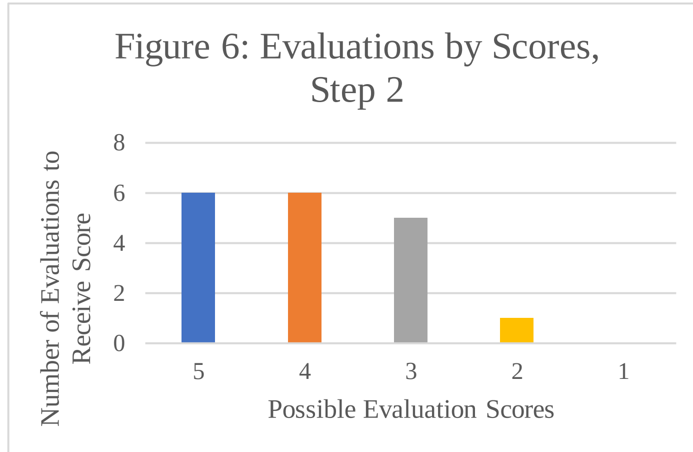
(Schwartz and Mayne 2005)

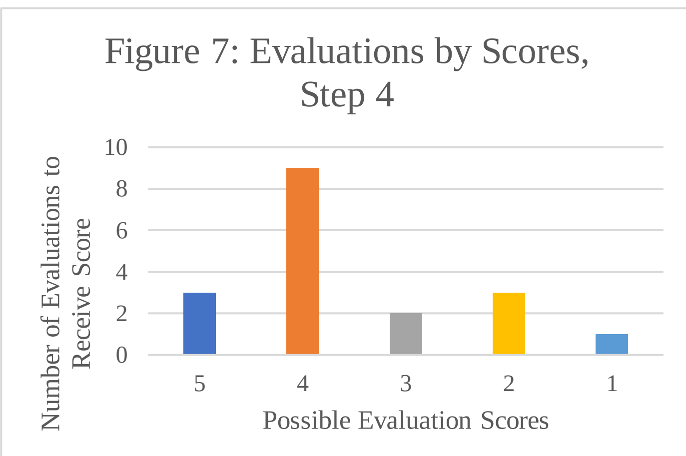
Table 4 – Description & lettering code of evaluation documents included in case study

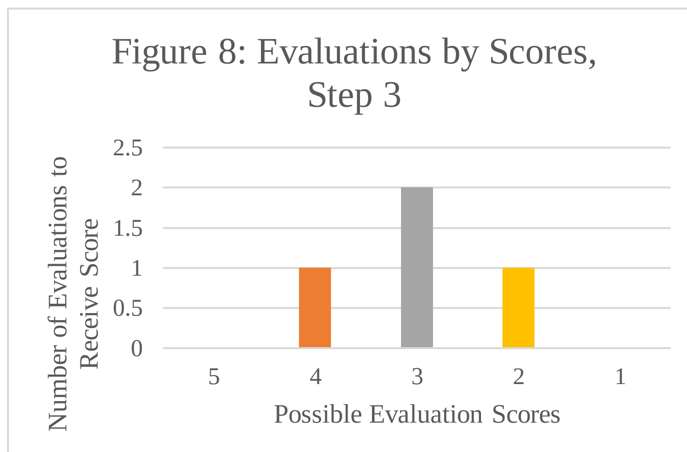
Compact Name	Existing Evaluations	Evaluation Type, Year	Identification Letter
Honduras Compact	Farmer Training & Development	Performance, 2012	A
	Transportation & Farm-to-Market Roads	Multiple, 2014	B
	Agricultural Public Goods Grant Facility	Performance, 2010	C
Nicaragua Compact	Rural Business Development Services	Impact, 2012	D
	Farming & Forestry	Performance, 2012	E
	ProNicaragua	Performance, 2012	F
	Property Regularization	Performance, 2012	G
	Rice & Plantain Farmers	Impact, 2015	H
	Transportation	Impact, 2014	I
El Salvador Compact	Community Services – Water & Sanitation	Impact, 2014	J
	Production & Business Services – Phases I & II	Multiple, 2013	K
	Investment Support & Financial Services	Performance, 2015	L
	Formal Technical Education	Multiple, 2016	M
	Non-Formal Skills Development	Performance, 2015	N
	Rural Electrification – Solar Panels	Performance, 2017	O
	Northern Transnational Highway	Impact, 2016	P
	Rural Electrification	Impact, 2016	Q
	Production & Business Services – Handicraft Value Chain	Impact, 2014	R

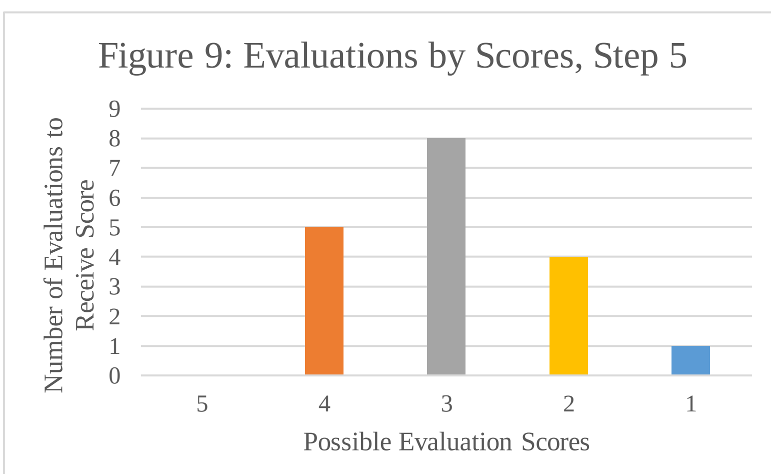
Figures 5-9 – Distribution of evaluation scores according to each step of document analysis











Appendix 1 – Scoring Rubric

	Excellent (5)	Good (4)	Fair (3)	Poor (2)	Bad (1)
Step 1 Concurrency of original project goals and evaluations findings	As compared to other evaluations in the case study, excellent evaluations structure findings based entirely on the original goals of the evaluated project. Evaluation goals and project goals align completely, and the evaluator has an excellent understanding of the intricacies of the project he/she is evaluating.	Good evaluations are those that structure findings mostly according to the goals of the evaluated project, although the evaluator might take away or add certain components. The evaluator has a good understanding of project goals, and most evaluation goals reflect those of the project. Certain original goals might be excluded from evaluation findings because of technical difficulties in the evaluation process.	Fair evaluations are those that might attempt to structure findings according to the goals of the evaluated project but fail to do so in some manner. The evaluator may have some misunderstanding of the original project goals. Several original goals are excluded from the evaluation findings because of technical difficulties or misunderstandings.	Poor evaluations are those whose findings are largely incongruent with the goals of the original project. The evaluator does not have a good understanding of original project goals and operates separately from project implementation personnel. Findings exclude many original project goals and instead only present that which the evaluator was able to establish according to his/her limited understanding of the project.	Bad evaluations are those who structure evaluations in a way divergent from the goals of the evaluated project. The evaluator is completely separate from the project and does not understand its original goals. Findings exclude almost all of original project goals and instead only reflect the narrow perceptions of the evaluator.
Step 2 Consistency of metrics used throughout evaluation	As compared to other evaluations in the case study, excellent evaluations clearly state metrics, use these metrics throughout the evaluation, and then present findings according to these metrics.	Good evaluations state metrics early on in the evaluation, although they could be further developed. These metrics are used throughout the evaluation but are not always clear to the reader. Findings are presented according to these metrics but might stray slightly from that which was originally intended.	Fair evaluations state metrics, but these metrics are not very clear to the reader. The evaluator uses these metrics – and others – throughout the evaluation. Their use is not entirely clear to the reader. Findings are presented using some of the original metrics but are not entirely consistent with original goals.	Poor evaluations may state metrics in a vague manner. The evaluator does not necessarily use these metrics throughout the evaluation, and the reader struggles in determining the metrics used. Findings are presented in a way not consistent with originally presented goals.	Bad evaluations may not state metrics at all. The evaluator uses haphazard metrics throughout the evaluation, and it is not apparent that the evaluator understands the intricacies of the project he/she is evaluating. Findings, although stated, do not relate at all to original statements.
Step 3 Concurrency of evaluation findings with findings of available secondary evaluations	Given the availability of secondary evaluations, excellent evaluations contain findings that are completely or almost-completely concurrent with those of secondary evaluations.	Good evaluations contain findings that are substantially concurrent with those of secondary evaluations, although one or two findings may slightly differ.	Fair evaluations contain findings that are concurrent with those of secondary evaluations, although several findings may notably differ.	Poor evaluations contain findings that concur – to a certain degree – with those of secondary evaluations, although many findings may differ substantially.	Bad evaluations contain findings that are differ greatly from those of secondary evaluations, although some basic findings might be concurrent.
Step 4 Quality of structure and findings of evaluation	As compared to other evaluations in the case study, excellent evaluations are those with the clearest structure, are the most user-friendly, and have findings that are clearly stated.	Good evaluations are those with a structure, user-friendliness, and description of findings that are relatively clear. Each could be better-defined by the evaluator, and the	Fair evaluations are those neither outstanding in structure nor poor in structure. Structure, although defined, is vague. The reader is left confused in several sections. Results, although	Poor evaluations are those below average. The evaluator has not structured the evaluation purposely and the reader is left confused in the majority of sections. Project results are not clear and	Bad evaluations are the worst in structure, user-friendliness, and presentation of results. The evaluator does not have a good understanding of the project, which reflects in a haphazard structure and

		reader is left confused in a few sections.	stated, need further development.	were not well-developed.	presentation of findings. The reader is left confused in nearly all sections.
Step 5 Extent to which project beneficiaries were considered in the evaluation process	As compared to other evaluations in the case study, excellent evaluations are both structured and worded in a way that, while satisfying the needs of donors, particularly stresses the importance of project impact on beneficiaries. Evaluation findings prioritize learning and present the impact the project had on beneficiaries. Community stakeholders and project beneficiaries are actively included in the evaluations process.	Good evaluations pay attention to project impact on beneficiaries but could also do so in a more meaningful way. A large focus of the structure and wording of these evaluations is to satisfy donor entities. Evaluation findings explain project impact on beneficiaries and learning, but these may not be the focus. Community stakeholders and project beneficiaries are included in important parts of evaluations process.	Fair evaluations mention project impact on beneficiaries but do not stress this in analysis nor in findings. These evaluations are worded and structured to benefit donor entities. Evaluation findings may touch on project impact on beneficiaries, but this is far from the focus of the findings portion of the evaluation. Community stakeholders and project beneficiaries are somewhat included in the evaluations process.	Poor evaluations give very little considering to project impact on beneficiaries and instead are almost entirely donor centric. These evaluations are worded and structured according to the desires of donor entities. Evaluation findings may mention project impact and learning in a cursory manner, but the priority of findings are those of the donors. Community stakeholders and project beneficiaries are not directly consulted.	Bad evaluations give no consideration to project impact on beneficiaries. These evaluations are worded and structured to be relevant only to donor entities and do not consider community stakeholders nor project beneficiaries. Evaluation findings do not mention project impact and pay only slight attention to learning. Findings are structures according to donor priorities.

Appendix 2 – Interview Instrument (with English & Spanish Questions)

Consent Script:

Hello, my name is Savannah Norman. Thank you so much for agreeing to participate today. As I've previously mentioned, I am a senior at Duke University writing my Public Policy & Spanish honors thesis on the evaluation practices of the Millennium Challenge Corporation's Central American compact projects. I am interested in interviewing the authors of the evaluations, such as yourself, to learn more about the evaluations process and ways in which it might be improved. Today's interview will take no more than 45 minutes, during which I will ask for your frank expert opinion. Are you willing to be interviewed?

Thank you.

I would like to record this interview, so that I can easily refer back to it. This recording is only meant only to expedite my academic process and will be used only for my research purposes. Do you give me permission to record?

Great. I will just double-check the audio and start the recording.

My Rationale:

I am interested in going into international development work and have academic and personal interests in Latin America. Studying the MCC's evaluations process has proven to be a great mix of my two interests – looking at both the project and policy side (for my PubPol major) and the more-qualitative side regarding the region and project beneficiaries.

Module 1 – The Evaluations Process

1. How did you come to be involved in evaluations process for the [name of project]? (¿Cómo llegó a ser involucrado con el proceso de evaluación para el proyecto de electrificación rural?)

- **Intent:** To learn about the way in which the evaluator came to be involved in the project. Obtain information about whether they regularly work with the MCC, or with projects in the region, or whether evaluators merely accept jobs as they come up.
  <u>Probes:</u>
  - Contracting process for evaluators
  - Potential interest in the region
  - Consistent evaluation partnership with the MCC

2. Did you receive any guidelines from the MCC regarding evaluation structure or content? (Recibió algún tipo de normas o alguna guía del MCC con respecto a la estructura o el contenido da la evaluación?)

- **Intent:** Reveal the extent to which the MCC seeks to establish consistency across evaluations. This will attempt to explain the differences I will encounter across the evaluations I will be studying.

Probes:
- Any attempt to establishing uniformity across evaluations?
- General requirements for evaluation

3. To what extent did you feel independent in conducting your evaluation? (¿Hasta qué punto se sintió independiente del MCC al realizar la evaluación?)

- **Intent:** Partial follow-up from previous question. Establish the extent to which evaluations feel independent in conducting their evaluations. Answers will have important implications for the integrity of the results of the evaluations.
  Probes:
  - Any pressure to publish positive findings?
  - Oversight by MCC staff members?

4. Despite the obvious difference in purpose between the performance and impact evaluations, what do you think the strengths and weaknesses are of the two different evaluation varieties?

- **Intent:** Only for those evaluators who conducted both impact and performance evaluations. Draw out the strengths and weaknesses of the two different evaluation types.
  Probes:
  - Qualitative vs. quantitative findings?
  - Cost of evaluations?

Module 2 – Project-Specific Questions

1. Can you describe any difficulties you faced while conducting these evaluations? (¿Puede describir algunas dificultades con las que se enfrentó al hacer esta evaluación?)

- **Intent:** Reveal any chronic difficulties or flaws in the data collection/evaluation development process. Can again speak to the quality of evaluations and the intricacies of project-specific evaluations.
  Probes:
  - Data collection
  - Establishing counterfactuals (if for impact report)
  - Measuring impact
  - Drawing conclusions

2. Can you describe your experience in or knowledge of [name of country] and/or Central America? Your experience with the subject matter of the projects? (¿Puede describir su experiencia en o su conocimiento de [nombre del país] y/o Centroamérica? Y también su experiencia con el [tema del proyecto]?)

- **Intent:** Obtain specifics about whether the evaluator is experienced in the region they are evaluating. Reveal extent to which evaluator considered culture, regional specificities, and values important to beneficiaries in their evaluation process.

3. Can you describe the extent to which stakeholders and project beneficiaries were involved in your evaluations process? (¿Puede describir el grado en el que los stakeholders (las partes interesadas) del proyecto o los beneficiarios del proyecto fueron incluidos en el proceso de evaluación?)

- **Intent:** Reveal extent to which stakeholders and beneficiaries were consulted. Obtain information about whose opinions were sought out, whose were not. Attempt to demonstrate whose opinions and priorities are being represented in these evaluations.
  Probes:
  - Indigenous People
  - Community Members
  - Individuals displaced by elements of project (roads, etc.)
  - Direct project beneficiaries, control groups

Module 3 – Learning & Growth

1. Having worked on large-scale, multi-year evaluations such as these, in what ways do you believe the evaluations process could be improved? (¿Después de haber trabajado en este proceso de evaluación, de qué manera cree que se podría mejorar el proceso de evaluación?)

- **Intent:** Obtain information about ways in which the evaluation process could be improved. Reveal potential flaws in the process not discernable from document analysis, or to reinforce trends.
  Probes:
  - More transparency
  - Involvement of stakeholders
  - Structural changes

2. How would you describe the impact of your evaluations, and of evaluations in general, for future development projects in Central America? (¿Cómo describiría el impacto y la importancia de esta evaluación – y de las evaluaciones en general – para futuros proyectos de desarrollo en Centroamérica?)

- **Intent:** Reveal steps for moving forward and potential answers to my second research question. Obtain information about whether evaluator believes their work is important and impactful.

3. How would you describe the extent to which these evaluations assessed the main goal of the MCC, the reduction of poverty through economic growth? (¿Cómo describiría el grado en el que esta evaluación abordó el meta principal del MCC, la reducción de la pobreza a través del crecimiento económico?)

- **Intent:** Reveal the perceived overall purpose of evaluations – whether they are conducted to benefit the bottom line of the MCC and donors, or whether they are conducted to actually assess the impact of MCC interventions on intended beneficiaries.

4. Do you think that consumption and income are the best measures to determine wellbeing/quality of life? (¿Cree que indicadores como el consumo y el ingreso son las mejores medidas para determinar el bienestar o la calidad de vida?)
Intent: Reveal whether evaluators value and/or see the potential of other, more holistic metrics. Reveal the potential weaknesses or shortcomings of metrics such as consumption and income.

Cool-Down Questions:
Do you have any other thoughts about the evaluation process or the specific projects that you would like to share?

Do you have any questions for me?

References

Actis Di Pasquale, E. (2008). La operacionalización del concepto de Bienestar Social: un análisis comparado de distintas mediciones. *Observatorio Laboral Revista Venezolana, 1*(2), 17-42.

Archibald, T., et al. (2018). "Every Practitioner a “Knowledge Worker”: Promoting Evaluative Thinking to Enhance Learning and Adaptive Management in International Development." New Directions for Evaluation **2018**(158): 73-91.

Anuario Estadístico de América Latina y el Caribe. (2019). CEPAL. Retrieved from https://www.cepal.org/es/publicaciones/ae

Bamberger, M. (2000). "The Evaluation of International Development Programs: A View from the Front." American Journal of Evaluation **21**(1): 95-102.

BenYishay, A. B. and R. Tunstall (2011). "Impact evaluation of infrastructure investments: The experience of the Millennium challenge corporation." Journal of Development Effectiveness **3**(1): 103-130.

Black, Maggie, and Maggie Black 1945. International Development: Illusions and Realities. New Internationalist, Oxford, UK, 2015.

Blandón, M. T. (2001). ¿ Cuál desarrollo para Centroamérica? *Encuentro*(56), 101-115.

Brown, N. M. (2019). Millennium Challenge Corporation: Overview and Issues.

Chac, M. (2008). Gobernanza y participación ciudadana en las políticas públicas frente al reto del desarrollo. *Política y cultura*(30), 9-37.

Chaguaceda, A. (2015). Regímenes políticos y procesos desdemocratizadores en Nicaragua y Venezuela. *Perfiles latinoamericanos, 23*(45), 5-29.

Chouinard, J. A., & Hopson, R. (2015). A critical exploration of culture in international development evaluation. *Canadian Journal of Program Evaluation, 30*(3).

Conceição, P. (2019). *Informe sobre Desarrollo Humano 2019*

Costa, S. (2012). Desigualdades, interdependências e afrodescendentes na América Latina. *Tempo Social, 24*(2), 123-145.

Delgado, L., Bachmann, P., & Oñate, B. (2007). Gobernanza ambiental: una estrategia orientada al desarrollo sustentable local a través de la participación ciudadana.

Deruyttere, A. (1997). *Pueblos indígenas, globalización y desarrollo con identidad: algunas reflexiones de estrategia*. Retrieved from Inter-American Development Bank

Flores, M. (2007). La identidad cultural del territorio como base de una estrategia de desarrollo sostenible. *Opera (Bogotá, Colombia), 7*(7), 35.

Freire, G., Díaz-Bonilla, C., Schwartz Orellana, S., López Soler, J., & Carbonar, F. (2018). *Afrodescendientes en Latinoamérica Hacia un marco de inclusión*. Grupo Banco Mundial.

GIEI Nicaragua. (2018). *Informe sobre los hechos de violencia ocurridos entre el 18 de abril y el 30 de mayo de 2018*. Retrieved from GIEI Nicaragua

Grasso, P. G. (2003). "What Makes an Evaluation Useful? Reflections from Experience in Large Organizations." American Journal of Evaluation **24**(4): 507-514.

Guerra-López, I. and K. Hicks (2015). "The participatory design of a performance oriented monitoring and evaluation system in an international development environment." Evaluation and Program Planning **48**: 21-30.

Guillen, A., Badii, M. H., Blanco, M., & Saenz, K. (2008). La participacion ciudadana en el contexto de desarrollo sustentable. *Innovaciones de Negocios, 5*(1), 131-146. Retrieved

Harvey, D. (2007). "A Brief History of Neoliberalism." Oxford University Press.

Hayes, N., et al. (2018). "Institutionalizing Inequality: Calculative Practices and Regimes of Inequality in International Development." Organization Studies **39**(9): 1203-1226.

Herrera, O. S. (2018). Golpe de estado, prensa escrita y asesinatos de periodistas en Honduras. *Revista de Ciencias Sociales*(162), 91-105.

Irarrázaval, I. (2005). *Participación ciudadana en programas de reducción de la pobreza en América Latina.* Paper presented at the documento presentado en el x Congreso Internacional del clad sobre la Reforma del Estado y de la Administración Pública, Santiago de Chile.

Juárez-Ramírez, C., Márquez-Serrano, M., Salgado de Snyder, N., Pelcastre-Villafuerte, B. E., Ruelas-González, M. G., & Reyes-Morales, H. (2014). La desigualdad en salud de grupos vulnerables de México: adultos mayores, indígenas y migrantes. *Revista Panamericana de Salud Pública, 35*, 284-290.

Kliksberg, B. (1998). Seis tesis no convencionales sobre participación. *Revista instituciones y desarrollo, 2*.

Lang, M. and D. Mokrani (2013). "Beyond Development, Alternative Visions from Latin America." Transnational Institute/Rosa Luxemburg Foundation.

Lehoucq, F. (2012). La economía política de la desigualdad en Centroamérica. *Anuario de Estudios Centroamericanos*, 79-108.

León Correa, F. J. (2011). Pobreza, vulnerabilidad y calidad de vida en América Latina: Retos para la bioética. *Acta bioethica, 17*(1), 19-29.

Leyva-Flores, R., Infante-Xibille, C., Gutiérrez, J. P., & Quintino-Pérez, F. (2013). Inequidad persistente en salud y acceso a los servicios para los pueblos indígenas de México, 2006-2012. *salud pública de méxico, 55*, S123-S128.

Más allá del ingreso, más allá de los promedios, más allá del presente: Desigualdades del desarrollo humano en el siglo XXI. (2019). Retrieved from Programa de las Naciones Unidas para el Desarrollo

Meyer, P. (2018). U.S. Foreign Assistance to Latin America and the Caribbean: FY2018 Appropriations.

Meyer, P. and E. Gracia (2019). U.S. Foreign Assistance to Latin America and the Caribbean: FY2019 Appropriations.

Millennium Challenge Corporation. "Independent Evaluations." https://www.mcc.gov/our-impact/independent-evaluations.

Millennium Challenge Corporation. "Where We Work" Retrieved from https://www.mcc.gov/where-we-work

Millennium Challenge Corporation. (2009). Congressional Notification [Press release]

Millennium Challenge Corporation. (2020). Press Release: MCC Marks Closeout of $277 Million El Salvador Investment Compact [Press release]

Mowforth, M. (2014). "The Violence of Development Resource Depletion, Environmental Crises and Human Rights Abuses in Central America." Pluto Press.

Néstor Feinstein, O. (2012). "Evaluation as a learning tool." New Directions for Evaluation **2012**(134): 103-112.

Öhler, H., et al. (2012). "Does conditionality work? A test for an innovative US aid scheme." European Economic Review **56**(1): 138-153.

Palomino Villavicencio, B., & López Pardo, G. (1999). Reflexiones sobre la calidad de vida y el desarrollo. *Región y Sociedad. Revista de El colegio de Sonora, 11*(17), 171-185.

Pena-Trapero, B. (2009). La medición del bienestar social: una revisión crítica. *Estudios de Economía aplicada, 27*(2), 299-324.

García Pinzón, V., & Rojas Ospina, E. J. (2020). La política de seguridad en El Salvador: la construcción del enemigo y sus efectos en la violencia y el orden social. *Revista de Estudios Sociales*(73), 96-108.

Quijano, A. (2000). Colonialidad del poder, eurocentrismo y América Latina. In: Clacso Buenos Aires.

Razo-González, A. M., Díaz-Castillo, R., Morales-Rossell, R., & Cerda-Barceló, R. (2014). Metaanálisis del concepto de calidad de vida en América Latina. Una nueva propuesta: sentido de vida Meta-analysis of the concept of life quality in Latin America. A new proposal: life meaning. *Revista CONAMED, 19*(4), 149-156.

Rivero, J. (2000). Reforma y desigualdad educativa en América Latina. *Revista Iberoamericana de educación*, *23*, 103-133.

Romero Cevallos, R. R. (2005). *¿CULTURA Y DESARROLLO? ¿DESARROLLO Y CULTURA? Propuestas para un debate abierto*. Retrieved from UNESCO Peru, UNDP Lima

Serrano, C. (1998). Participación social y ciudadanía. Un debate del Chile contemporáneo. *Santiago de Chile: División de Estudios del Ministerio de Planificación, Gobierno de Chile*.

Schwartz, R. and J. Mayne (2005). "Assuring the quality of evaluative information: theory and practice." Evaluation and Program Planning **28**(1): 1-14.

Sturdy, J., et al. (2014). "Learning from evaluation at the Millennium Challenge Corporation." Journal of Development Effectiveness **6**(4): 436-450.

Telles, E., & Flores, R. (2013). Not just color: Whiteness, nation, and status in Latin America. *Hispanic American Historical Review*, *93*(3), 411-449.

Villalba-Eguiluz, C. U. and I. Etxano (2017). "Buen Vivir vs Development (II): The Limits of (Neo-)Extractivism." Ecological Economics **138**: 1-11.

Villatoro, P. (2012). La medición del bienestar a través de indicadores subjetivos: una revisión.