What makes a good project?
Success factors of the World Bank education development projects

By
Bommi Lee

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirement
for the degree of

DOCTOR OF PHILOSOPHY
in
Leadership and Policy Studies

May, 2016

Nashville, Tennessee

Approved:
Professor Stephen Heyneman
Professor Will Doyle
Professor Ron Zimmer

To my parents and my husband with love and gratitude
감사합니다. 그리고 사랑합니다.

ACKNOWLEDGEMENTS

I am grateful to everyone who has helped me sustain through this whole process. Words are not enough to show how thankful I am for their support.

I am very grateful to Dr. Stephen Heyneman, Chair of this dissertation and my academic advisor, for encouraging me to pursue my study in the field of international development and education as well as providing his keen insights and guidance throughout my journey. Each member of my Dissertation Committee has provided me with extensive guidance on the topic and helped me with practical skills that I needed to deal with the complex questions in this study. Special thanks to Dr. Birger Fredriksen, who actively helped me in recruiting the interviewees, as well as provided detailed comments on every page of my dissertation.

Many thanks to Mr. Sverrir Sigurdsson who also helped me in recruiting the interviewees, Mr. DaeIn Chong who gave me general guidance on IEG evaluation ratings and answered all my trivial questions about using this as data. This study would not have been possible without the participation of the interviewees who shared their personal experiences. I would also like to thank Dr. Christopher Kilby and Dr. Matthew Winters for giving me technical help with the IEG ratings data.

Nobody has been more important to me in the pursuit of this study than my family. I would like to thank my parents and my sister whose love and support are endless with whatever I pursue. Most importantly, I thank my beloved and infinitely supportive husband, Jaehoon Kim, and my amazing son, Sean Kim, who provide unending inspiration.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Canadian International Development Agency (CIDA)

Country Policy and Institutional Assessment (CPIA)

Craig Burnside and David Dollar (BD)

Development Assistance Committee (DAC)

East Asia and the Pacific (EAP)

Education for All (EFA)

Europe and Central Asia (ECA)

Evaluation Summaries (ES)

Fast Track Initiative (FTI)

Generalized Method of Moments (GMM)

Global Partnership for Education (GPE)

Government Effectiveness (GE)

Gross Domestic Product (GDP)

Gross National Income (GNI)

Hierarchical Linear model (HLM)

Implementation Completion Report (ICR)

Implementation Status and Results Report (ISR)

International Bank for Reconstruction and Development (IBRD)

International Development Association (IDA)

International Monetary Fund (IMF)

Latin America and Caribbean (LCR)

Middle East and North Africa (MNA)

Millennium Development Goals (MDGs)

Ministry of Education (MOE)

Missing At Random (MAR)

Official Development Assistance (ODA)

Operations Evaluation Department (OED)

Ordinary Least Square (OLS)

Organisation for Economic Co-operation and Development (OECD)

Project Appraisal Documents (PAD)

Project Completion Reports (PCR)

Project Development Objective (PDO)

Project Performance Assessment Reports (PPAR)

Project Performance Audit Reports (PPAR)

Receiver Operator Characteristic (ROC)

South Asia (SAR)

Staff Appraisal Report (SAR)

Sub-Saharan Africa (AFR)

United Nations (UN)

United Nations Educational, Scientific and Cultural Organization (UNESCO)

United Nations Security Council (UNSC)

United States Agency for International Development (USAID)

Universal Primary Education (UPE)

World Bank Independent Evaluation Group (IEG)

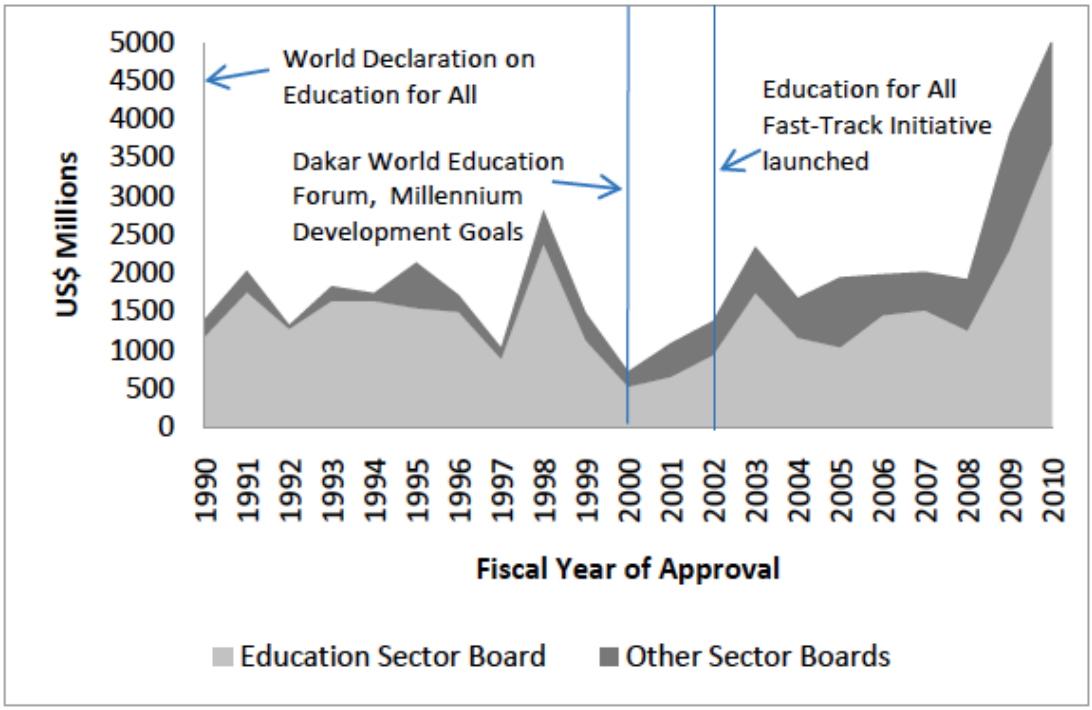
World Development Indicators (WDI)

Worldwide Governance Indicators (WGI)

CHAPTER I

INTRODUCTION

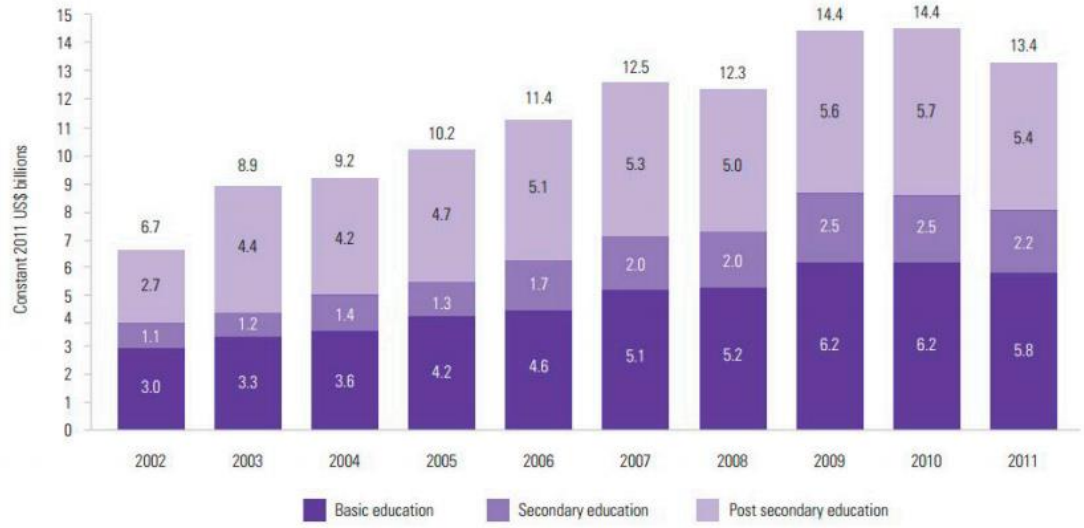
Foreign aid has continuously increased since the 1960s. Aid increased substantially since the 2000s when the Millennium Development Goals (MDGs) were established. Developed countries doubled their annual aid flows, and tripled aid specifically for Africa (Birdsall & Savedoff, 2010). By 2012, the foreign aid amounted to more than $125 billion a year (OECD, 2015). Charitable foundations and philanthropic international programs that actively participate in the international donor community also increased substantially, although the size of their commitments is still very small compared to that of official development assistance from major donor countries (Birdsall & Savedoff, 2010; Heyneman & Lee, 2013). Aid to the education sector also expanded, mirroring the increase in overall aid levels (UNESCO, 2013b). Both the Education MDGs and the six EFA (Education for All) goals were established in 2000 and targeted at 2015 (Burnett & Felsman, 2012). The World Bank and other development partners launched the Education for All Fast Track Initiative (FTI) in 2002 to meet the educational goals of EFA and MDGs. The World Bank committed about $23 billion to finance education programs between 2001 through mid-2010 (World Bank IEG, 2011). As illustrated in Figure 1, new education commitments dropped to their lowest since the financial crisis in the late 1990s, however, the amount doubled from $1 billion annually in 2000 to $2 billion annually by 2007–08 (World Bank IEG, 2011).



Source: World Bank data.
Note: Financial commitments for the part of projects that represents education are attributed to the year of approval. Additional financing is attributed to the year that the additional financing was approved. Figures are in nominal US dollars.

Figure 1 New World Bank commitments to education, 1990-2010 ($ millions)

However, an increase of aid in the education sector declined after 2010 (Heyneman & Lee, 2013; Mundy, 2006; UNESCO, 2013a). The share of official aid in the education sector has been around 13% over the past decade, but there was about a 7% reduction in education aid between 2010 and 2011, while there was only a 3% reduction in total aid over the period (see Figure 2) (UNESCO, 2013a). The World Bank's new commitments to education also stopped increasing after the most recent financial crisis (World Bank IEG, 2011).



Source: EFA Global Monitoring Report team analysis based on OECD Creditor Reporting System (2013).

Figure 2 Total aid disbursements to education, 2002–2011

The recent decline in education aid along with overall aid level indicates that aid budgets are coming under pressure as donor countries find it difficult to manage unprecedented levels of debt and unsustainable budget deficits (Overseas Development Institute (ODI), 2009). Thus there has never been greater need to make the case that aid is used effectively to achieve results (Bermingham, Christensen, & Mahn, 2009). However, the aid effectiveness literature showed insufficient evidence to support that aid has a positive effect. Some studies show that foreign aid has a positive effect on economic growth under good policies and institutions (Burnside & Dollar, 1997; Svensson, 1999; Burnside & Dollar, 2004; Collier & Dollar, 2004), whereas others show no effect of foreign aid (Boone, 1994; Dalgaard & Hansen, 2001; Easterly, Levine, & Roodman, 2004; Rajan & Subramanian, 2005). As education plays a fundamental role in poverty alleviation and economic growth (Hanushek & Woessmann, 2008; World Bank

Independent Evaluation Group, 2011), some recent studies examined the effectiveness of educational aid. Aggregate level studies show a moderately positive effect of educational aid on school enrollment and completion rates, though the effects are not robust across different model specifications (Michaelowa & Weber, 2006; Michaelowa & Weber, 2008; Dreher, Nunnenkamp, & Thiele, 2008; Christensen, Homer, & Nielson, 2011). Despite the skeptical results of aid effectiveness, the international donor community, particularly at the point when the MDG agenda (2000–2015) is reaching its end, became aware of reduced poverty during the MDG period. The extensive ongoing debate on the Post-2015 Development Agenda is quite favorable to aid the recent "Report of the High-Level Panel of Eminent Persons on the Post-2015 Development Agenda," which acknowledges the large effort by donors that contributed to remarkable achievements after the MDGs; half a billion fewer people are living below the international poverty line of $1.25 a day, child death rates have fallen by more than 30%, with about three million children's lives saved each year compared to 2000, 590 million children in developing countries – record number– attended primary school. The panel suggests that such success calls for more action (United Nations, 2013).

Not only is there optimism on aid effectiveness by the international community, but there is also another reason for aid effectiveness to be observed. Foreign aid is driven by the donor's political and long-term strategic interests as well as developmental and humanitarian goals to promote long term economic growth and political stability (Akhand & Gupta, 2002). Because of this political nature of foreign aid, aid is likely to continue, albeit perhaps on a declining basis and taking different forms regardless of some empirical studies questioning the impact.

Therefore, it is more appropriate to assume that foreign aid works and explore the conditions under which foreign aid has an effect. In fact, the overarching literature on foreign aid is policy debates about what 'matters' for development (Moloney, 2009). Also, since aggregate level studies have limitation in explaining the large variation in project success or failure (Doucouliagos & Paldam, 2009) and most foreign aid is delivered in the form of projects, studies need to examine the aid effectiveness at the project level. Until recently, foreign aid scholars have had limited opportunities to engage in sector-specific analysis and project level analysis due to data scarcity (Moloney, 2009). The World Bank project performance ratings data provides a direct input and output relationship of project characteristics and project outcomes. The World Bank is also the most influential multilateral organization in the education sector from the international donor community, as it has the largest funding source. Therefore, this study examines the determinants of the education project outcomes of the World Bank.

This study is divided into two parts. The quantitative part of this study aims to examine the education project characteristics significantly associated with project outcomes using the World Bank Independent Evaluation Group (IEG) project performance data. I use the education projects that were implemented between 1996 and 2011. The characteristics are divided into two: country level characteristics such as spending on education sector and GDP growth, and project level characteristics such as project size, duration, and type of loans. As one of the project level characteristics, I include a dummy variable for development policy loans, as there have been discussions on aid modalities, that channeling aid through the national budget might be more

effective since it strengthens ownership and national systems, as well as addresses inter-sectoral issues (Fredriksen, 2011a).

The first research question, therefore, is "What factors explain the variation in project performance of education development projects at the World Bank?" Using the data available from the IEG data and country level data from various sources, I apply a linear probability, logit, ordered logit, and logit models with country fixed effects to examine the determinants of project outcomes.

The qualitative part of the study aims to use the World Bank staff's experience and report their opinions on project determinants and outcomes. The IEG data does not contain detailed information on each project. For example, there is no information on the implementing staff's background and experience that might play a significant role in project's success. Not only the implementing staff's quality, but there could be many other factors that might be considered important for project outcomes as the Bank mechanism of project implementation is rather complicated. Therefore, the second research question is "What factors are perceived as important for World Bank education project outcomes?"

To answer the research question for the qualitative part, I collected interview data from 15 current and retired Bank staff. Semi-structured individual interview was conducted. All interviews are transcribed and coded using Nvivo. Results from the quantitative data combined with the qualitative data presents a larger picture of the significant factors for project outcomes.

The rest of this dissertation is organized into three sections. In the background, I explain that this study focuses on the relationship between project characteristics and

outcomes within the larger context of aid effectiveness literature. I present a typical Bank project cycle and explain what is being done at each phase of the cycle. I also describe how the Bank's lending policy has changed over time since the 1970s. In the literature review section, I present the findings from the most relevant studies. Since this is the first study to focus on education sector using project level data, I review studies from four different areas: 1) macro level studies on foreign aid in education from economics literature 2) studies that focus on World Bank projects using the project level data from economic literature 3) determinants of project outcomes from the project management literature and finally 4) studies that describe the Bank's education projects from educational management literature. In the research design section, I describe the data and the analytical strategy for both the quantitative and qualitative parts of this study.

Before moving on to the next section, it is crucial that I clarify the term "project success" used in this dissertation. Although the term "project success" sounds vague, in this study, I only refer to the short-term project outcomes, which is defined by project evaluation ratings. Although measuring the long-term outcomes of projects, which is the program impact for the beneficiaries, are the most ideal and useful for practitioners and policy makers in the field of international development, long-term project outcomes are not obtained for this study and thus not included in the data. Therefore, in this study, the term "project success" is referred to as short term project outcomes, outcomes that are measured/evaluated usually after two to three years of implementation completion.

CHAPTER II

BACKGROUND

1. Narrowing the scope from general aid to World Bank education projects

Foreign assistance was started after World War II to aid the war-affected countries for the purpose of reconstruction, extending political influence and altruism (Heyneman & Lee, 2013). It began when the U.S. transferred 2–3% of its national income to restore the European economic system (Tarp, 2010). This Marshall Plan, also known as the European Recovery Programme, was considered the most successful aid program. Thereafter, the focus of foreign aid shifted to the newly independent and less industrialized countries in Africa, Asia and Latin America. These countries were seen as strategically important countries in the emerging Cold War between the Soviet Union and the United States (Chabbott, 2003). Foreign aid programs offered technical aid to less industrialized countries. By the 1990s, most industrialized countries were contributing to multilateral development aid and all the OECD members were involved in bilateral development programs (Chabbott, 2003). An increasing number of non-governmental organizations also became involved in the development policy arenas. More recently, the countries that "graduated" from low-income countries, such as China, India and Brazil, became new donors (Mundy, 2006; Mundy, 2002).

Foreign aid is delivered through different types of channels. Bilateral official aid is channeled directly from the donor government agencies (e.g. United States Agency of

International Development, Canadian International Development Agency) to the recipient countries. Multilateral official aid is channeled through multilateral organizations such as the World Bank and UNESCO. United Nations organizations provide grants that do not have to be repaid, whereas the World Bank, the International Monetary Fund (IMF) and regional banks provide loans that need to be repaid (Heyneman & Lee, 2013). Charitable foundations (e.g. Save the Children, CARE) and religious philanthropies (e.g. World Vision) also actively participate in international educational development (Heyneman & Lee, 2013).

Among the different types of foreign aid delivery channels, I narrow the focus of this study to multilateral assistance. Bilateral assistance tends to reflect the political, economic and geographical interests of the donor governments. The spending of charitable foundations and religious philanthropies is difficult to track, as they are not obligated to report to the OECD on the amount of aid disbursed. On the other hand, the OECD keeps a database of all Development Assistance Committee (DAC) members and multilateral organizations' amount of Official Development Assistance (ODA) every year in their Creditor Reporting System. Among the multilateral organizations, the World Bank keeps a record of all implemented operational projects and the project performance ratings.

Moreover, within the larger context of international donor community, the World Bank has a substantial influence on the direction of donors' policies in the education sector. The Bank is the largest single international source of education finance, and in FY 2013, the Bank disbursed 2.9 billion dollars for education operations (Mundy, 2002; King, 2013). The magnitude of resources is much larger than other United Nations

institutions (Marshall, 2008). The World Bank also has several pooled trust funds for education in conflict-affected states. It is also the host of the Global Partnership for Education (formerly the Fast Track Initiative). The Bank is a core agency of global governance that plays a crucial role in the business of development (Marshall, 2008). Therefore, the long-term success of the aid effectiveness agenda will depend on changes in operational and management culture at the World Bank (Bermingham, Christensen, & Mahn, Aid effectiveness in education: Why it matters, 2009).

Within the World Bank, there are different types of financial services. There are lending instruments, grants, co-financing and trust funds (World Bank, 2007). There are two types of lending instruments: investment lending and development policy lending. Investment loans provide long-term financing for a range of activities and has, on average, accounted for 75–80 percent of all Bank lending (World Bank, 2007). Development policy loans, previously called adjustment loans, provide quick disbursement assistance to countries to support structural reforms. On average, this has accounted for 20–25% of total Bank lending (World Bank, 2007).

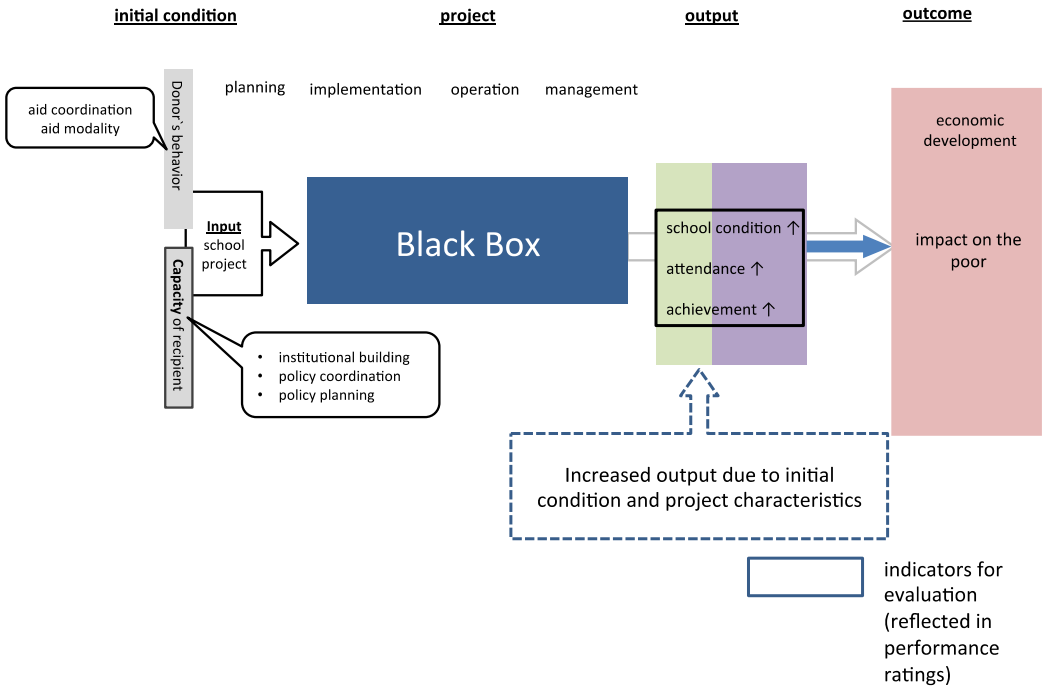
The Bank has a few grants that are funded directly or through partnerships. Grants are usually used as seed money for innovative projects. Co-financing is any arrangement in which the Bank's fund is combined with funds provided by sources from non-recipient countries. The Bank can co-finance projects in collaboration with other donor governments, multilateral institutions or commercial banks (World Bank, 2007). Trust funds are an arrangement between the Bank and a donor or a group of donors, in which the donor entrusts the Bank with funds used for specific development activity. These categories are not mutually exclusive. An investment lending, for example, can be co-

financed with other donors. In this study, based on data availability, I distinguish between investment loan and development policy loan.

In this dissertation, I focus on educational lending projects that are financed by the International Bank for Reconstruction and Development (IBRD) and International Development Association (IDA). IBRD is the original World Bank institution, and it provides combination of financial resources, knowledge and technical services, and strategic advice to developing countries, including middle income and credit-worthy lower income countries (World Bank, 2016a). The IBRD provides loans that charge interest and are medium-term (normally up to 10 years) (Gilbert & Vines, 2000). Although IBRD loans remain the Bank's main activity, the Bank has an aid agency, the IDA. The IDA is part of the World Bank that helps the poorest countries by providing loans (called "credits") and grants for programs to boost economic growth, reduce inequalities, and improve people's lives (World Bank, 2016b). IDA complements the World Bank's original lending arm, IBRD. IDA lends money on concessional terms, meaning that the IDA credits have zero or very low interest over 25 to 38 years. IDA also provides grants to countries at risk of debt distress (World Bank, 2016b). The IBRD and IDA share the same staff, report to same senior management, and use the same standards when evaluating projects (World Bank, 2007, p. 11).

This study also focuses on project level characteristics rather than macro economic factors. To understand aid effectiveness it is critical to examine the causality chain between foreign aid to final developmental outcomes (Bourguignon & Sundberg, 2007). Figure 3 shows that initial conditions of donors and recipients and inputs leads to changes in outcomes measured by indicators such as GDP growth and poverty rate.

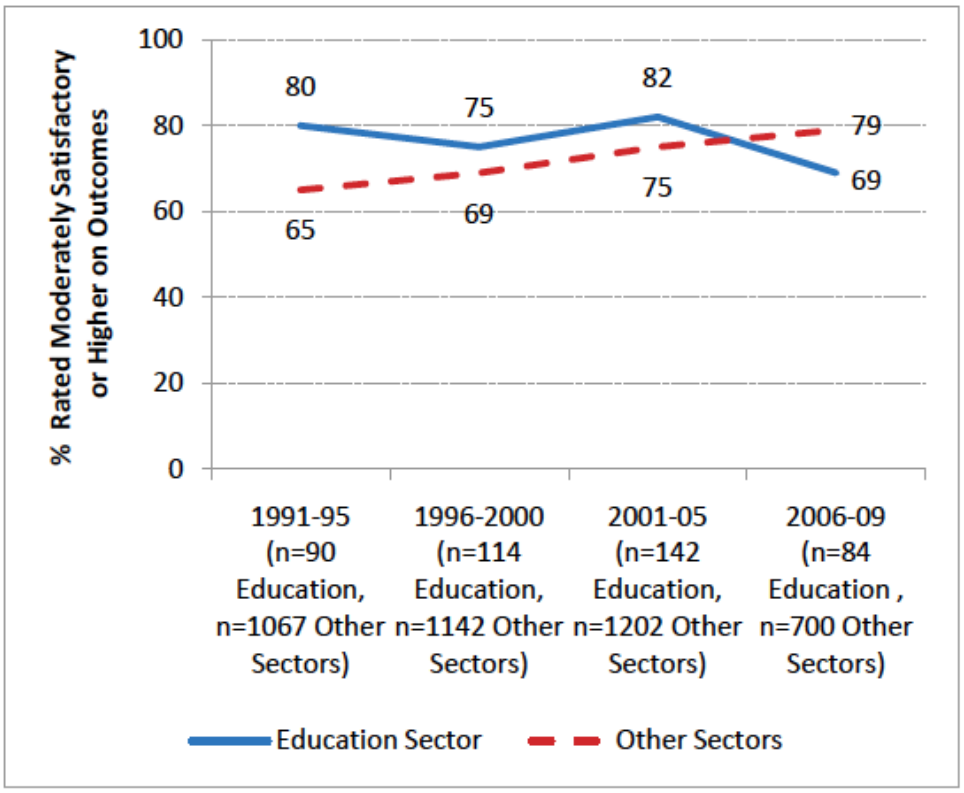
Previous studies on aid effectiveness have focused on the beginning and the end of this chain, rather than understanding the 'black box'. Standard cross-country regression analyses observed the aid-growth relationships using macro-level economic data (Nissanke, 2010). However, relating donor inputs and developmental outcomes directly through the black box will often lead nowhere (Bourguignon & Sundberg, 2007). Therefore, this study attempts to examine the relationship between the factors in the 'black box' and the project output. Several factors are inside the 'black box', such as project size, project duration and donor–recipient relations. In this study, the World Bank project performance rating is regarded as the project output.



Source: Adapted from Jerve and Nissanke (2007) and JBIC Secretariat. Modified by author.

Figure 3 Causality chain of development assistance projects

According to a World Bank Operations Evaluation Department (OED) report, the Bank project outcomes have improved a great deal over the past ten years: in 1995, 67 percent of the projects were rated satisfactory compared to 78% in 2004 (World Bank Operations Evaluation Department, 2005). Compared to the constant increase in the percent of projects rated satisfactory in other sectors, the ratings of education projects have declined recently (Figure 4). The IEG analyzes that the decline in education project performance is due to overambitious project goals, technical design issues, and external factors beyond the control of the project (World Bank Independent Evaluation Group, 2011, p. 35).

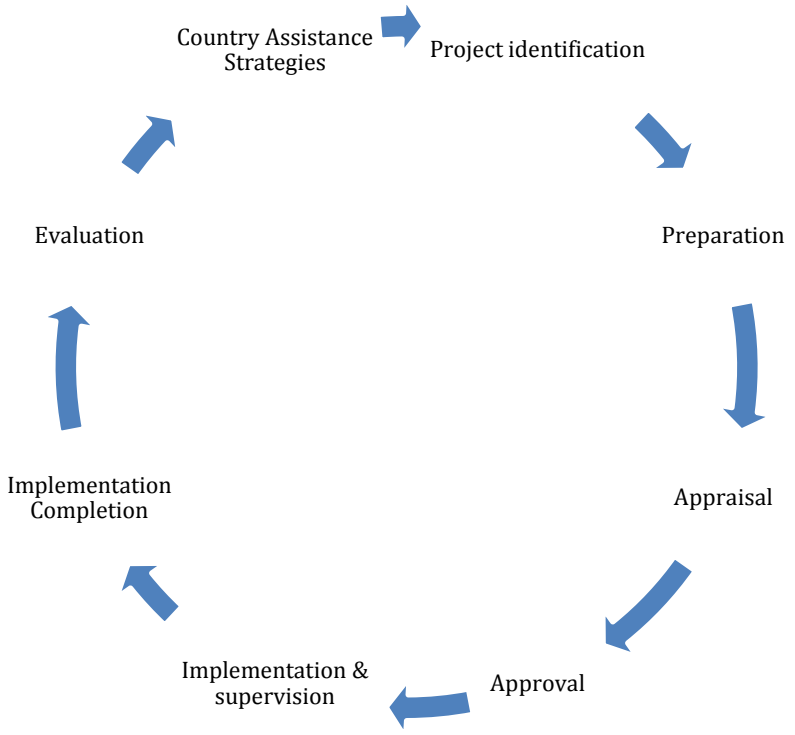


Source: World Bank data.
Note: The decline in Education Sector performance between 2001–05 and 2006–09 is statistically significant at p = 0.02.

Figure 4 Trends in performance of education projects and projects in other sectors, by fiscal year of exit, 1991–2009

2. Classic project cycle at the Bank

Understanding the project cycle at the Bank and examining each project phase is necessary to learn about the mechanism of project implementation and its outcomes[1]. Figure 5 illustrates a classic World Bank project cycle. In this section, I describe each phase of a classic project cycle and some issues that might arise at each phase. Other types of projects do not go through the same process; however, in this study I assume that each project goes through the classic project cycle.



Source: World Bank (2007, p. 77). Figure 2.5. Modified by author.

Figure 5 Project cycle at the World Bank

[1] I call output in Figure 3 as project outcomes in this study.

The planning phase: Identification, preparation, appraisal and approval

The project cycle can be largely divided into three phases: planning, implementation, and evaluation (Kilby, 2001). Planning consists of identification, preparation, appraisal, negotiation, and board approval. Each department within each region generates the ideas for what should be funded (Heyneman, 2003). The teams of the Bank work with the borrower government to identify projects that closely align with the borrower government's national priorities (World Bank, 2007). During the identification phase, the Bank develops a conceptual basis for the project and prepares a project concept note.

Once the objectives and concepts of the projects are set, the borrower country prepares the project. The government investigates technical, institutional, economic, environmental, and financial issues facing the project and prepares an assessment of projects for Bank financing (World Bank, 2007). The preparation phase can last from a few months to three years, depending on the complexity of the project (World Bank, 2007). This pre-appraisal period is one of the key stages of project cycle, as the first preliminary assessment of the quality and feasibility of projects are being conducted.

After the project preparation, the Bank staff evaluates the project plan by spending some time in the client country (World Bank, 2007). The staff presents the results of the plan evaluation in Project Appraisal Documents (PAD), formerly called Staff Appraisal Report (SAR), which is reviewed by the Bank management and financial team (World Bank, 2007). This document is also the starting point for negotiations with the borrower country (Kilby, World Bank borrower relations and project supervision, 2001). The document used in negotiations is called "Credit Agreement," which represents

the legal agreement setting out the conditions governing the credit (IDA) or loan (IBRD). During appraisal by World Bank staff, the project "may be extensively modified or redesigned" (Baum, 1982, p. 17). Changes are reflected in the project appraisal documents, and both the Bank and the borrower government agree on the terms and conditions of the loan (Kilby, 2001; World Bank, 2007).

Then the project documents are submitted to the board of directors of the Bank for approval (Kilby, 2001). When approved, the loan or credit becomes effective, and the document becomes the standard for performance evaluation (Kilby, 2001; World Bank, 2007). During the approval phase, some projects that fail to get approval are dropped[2].

Several factors may be associated with project outcomes during the stage of project planning. Deficient design and preparation leads to delays in implementation (Jones, 1992). Intangible or broad project objectives are often challenges for implementation (IEG, 2009). Over ambition followed by inadequate readiness for implementation are also known as factors associated with unsatisfactory performance of education projects (World Bank IEG, 2011).

Also, a number of studies examined whether political factors (e.g. the borrower country holds a non-permanent seat on the UN Security Council) accelerate the process of project preparation, increase the number of projects and the amount of loan approved. However, these studies did not examine impact on outcomes (Kilby, 2012; Dreher, Klasen, Vreeland, & Werker, 2010; Kaja & Werker, 2010; Andersen, Hansen, & Markussen, 2006). Only Kilby (2012) describes how political economy factors are exogenous to latent project quality and uses the duration of project preparation based on

[2] The World Bank keeps document of these projects and there are 70 projects that are "dropped" in education sector. "Dropped" projects do not necessarily mean that these projects are not approved by the Board.

the exogenous variation of political economy factors to look at the impact of preparation on outcomes.

The Bank culture can indicate potential influence on project characteristics or outcomes. There was competition among staff to be responsible for new loans between and within sectors, thus incentives for the staff to create new loans outweighed the incentives for implementing projects more effectively (Heyneman, 2003). Career prospects of Bank loan officers depended largely on how fast they have a project with sizable loans approved, and how quickly they have the project's disbursements implemented throughout the life of a project (Jones, World Bank financing of education: Lending, learning and development, 1992). This implies that the Bank culture did not necessarily motivate the Bank staff to achieve better project outcomes, but rather increased the number of loans.

The implementation and supervision phase: the black box

After board approval, a project enters its implementation phase. The borrower country implements the project. The implementation is divided into different funding phases, whose frequency and scale are determined during the preparation phase (Chauvet, Collier, & Duponchel, 2010). The World Bank conducts supervision for each project and depending on the conclusions from the supervision report, the funding can be aborted, or scaled down (Chauvet, Collier, & Duponchel, 2010). The project teams report on progress twice a year, but there is a trend toward continuous supervision by teams in the country (Marshall, 2008).

Bank management decides the effort put into supervision and funds are allocated for supervision. While monitoring is the main activity of supervision, it also includes advising on management issues and consulting technical assistance (Kilby, 2000). The Bank operations staff spends about an average of 12 weeks of supervision annually (Kilby, 2000). A good supervision usually involves close relationships with the borrowing institutions and a clear understanding of the appropriate role for the Bank (Marshall, 2008). As the project operation reaches its completion, the borrower and Bank staff prepares a completion note, which is reviewed by the IEG (Marshall, 2008).

Certain factors hamper the implementation of projects. Some factors are outside the Bank's control. For instance, the 1972 earthquake in Nicaragua devastated the project. In Sudan, the government frequently changed, which impeded the implementation of projects (Jones, 1992). However, some factors are under the Bank's control. One of these factors is the design of projects. A project with complex design not only delays implementation, but also makes the modification during implementation difficult (Jones, 1992).

The relationship between the Bank staff and the borrower government is another challenge for implementation. Each country works with a country team that consists of Bank staff and locally based staff. The team's collaboration varies from country to country. However, teams often work with contrasting, fragmented, and factious arrangements (Marshall, 2008).

The staff and the project team seem to be the keys to successful implementation of the project. The incentives provided to the Bank staff to work particularly in the social sector themes were important (Koeberle, Bedoya, Silarszky, & Verheyen, 2005).

Continuity in staffing is also important as the staff shift regions and offices from time to time. Also, a project team that has experience in the relevant country or sector would perform much better (Koeberle, Bedoya, Silarszky, & Verheyen, 2005).

Budget allocations also affect project performance. Even the most qualified staff members are hampered by insufficient budgets for implementation (Koeberle, Bedoya, Silarszky, & Verheyen, 2005). Inadequate funding and cost overruns lead to serious delays of implementation and non-completion of projects (Kaufmann & Wang, 1995). In fact, many projects suffered from cost overruns and delays in implementation (Jones, 1992). This is sometimes due to macroeconomic factors, such as large fiscal deficits or high inflation rates, making the local government incapable of disbursing the amount it promised in the preparation stage.

Koeberle, Bedoya, Silarszky, & Verheyen (2005) also report that supportive local environment is critical for a satisfactory outcome. Support by government leaders and public sector managers and their staff is important for successful project implementation. A well-structured decentralized management system was also a factor that contributed to the success of a project. The World Bank also listed some characteristics of highly satisfactory projects in the social sector, and many included local participation. A successful project was designed in which the beneficiaries participated in the design to help identify desired outcomes. Beneficiaries were also involved in project implementation, and their participation was culturally appropriate. A successful project was also built on local NGO capacity. All of these factors that might affect project success were not captured in the IEG database. However, it is important to understand the complexity of the implementation process.

The evaluation phase

When a project is completed, the IEG conducts an audit to measure the project's outcome against the original objectives. The IEG (formerly called OED) is a semi-autonomous organization under the World Bank Group that was established in 1973. The IEG reports directly to the Board and, in principle, its staff is not permitted to move to positions in other parts of the Bank. The outcomes of those projects are rated independently.

The IEG rates project outcomes according to three criteria: 1) relevance, which is the relevance of the objectives of the projects in regard to the country's needs and institutional priorities 2) efficacy, which measures the extent to which the objectives have been achieved and 3) efficiency, i.e. the extent of the achievement of the objectives without using more resources than necessary (cost/benefit analysis) (World Bank Independent Evaluation Group, 2009). Ideally, a satisfactory project fully delivers on its development objectives, appear likely to be sustained after the project is completed, and generate clear benefits in an efficient manner (World Bank Independent Evaluation Group, 2009). Based on these criteria, the outcome of the project is rated from 0 (highly unsatisfactory) to 5 (highly satisfactory). The rating scale changed from binary (satisfactory/unsatisfactory) to six scales in 1995.

The second research question arose from questioning the definition of each scale of project outcome. The definition of each scale for project outcome is not quite clear. The Operations Policy and Country Services ICR guideline documents define the term "shortcomings" as "may have to do with either the number of objectives that are not achieved or the extent to which one or more objectives are not achieved (or are not

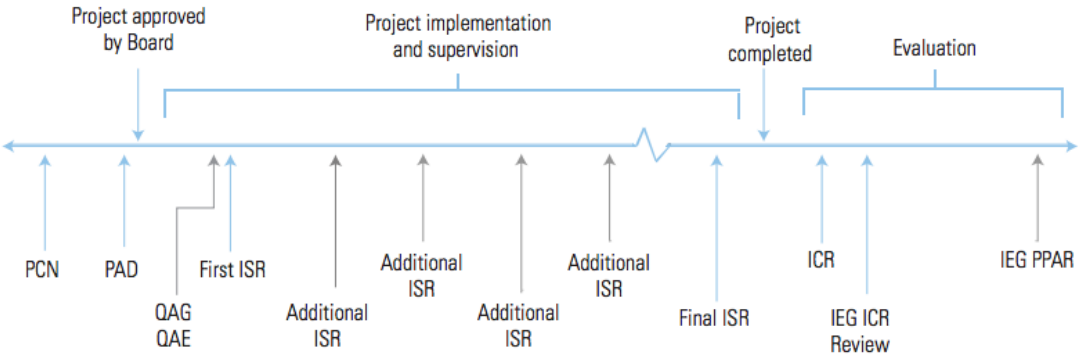
expected to be achieved)". This definition provides room for subjective judgment. Thus, the project performance ratings are likely to be judged by the individual evaluator who complete the Implementation Completion Report, and by the evaluator at IEG who reviews the ICRs.

Project evaluations began at the World Bank ever since the 1960s, when McNamara introduced the idea of ex post evaluation: assessment of a completed project or program (Xu & Weller, 2009). In 1973, the Operations Evaluation Department (OED) was created. The OED, now called Independent Evaluation Group (IEG), has its own budget, staff and functions and reports directly to the Board of Executive Directors (Xu & Weller, 2009; Marshall, 2008).

There are pros and cons of having the IEG within the Bank. Critics argue that the independence of the IEG is compromised when it remains part of the Bank and it is staffed by regular employees (Xu & Weller, 2009). Although the board is not part of the Bank management, since the board approves the IEG budget, the content and direction of work programs are strongly influenced by the Bank management through budgeting (Kopp, 2003).

On the other hand, supporters argue that keeping the IEG as part of the Bank is necessary for meaningful evaluation, because the mechanism at the Bank is complicated and it is not easy for outsiders to understand the mechanism and evaluate the projects. Also, evaluating a project requires access to detailed information from both operations and borrowing countries, and the operational staff is in control of information about the staff's own performance (Xu & Weller, 2009). Thus an insider would have better access

to the information than an outsider, and also an inside evaluator would be able to interact with the operational staff better than an outside evaluator.



Sources: Operations Policy and Country Services, World Bank; IEG.
Note: PCN = Project Concept Note; PAD = Project Appraisal Document; ISR = Implementation Status and Results Report; QAE = Quality at Entry; ICR = Implementation Completion Report; PPAR = Project Performance Assessment Report.

Figure 6 Key evaluation ratings

Figure 6[3] illustrates some key ratings done by the operational staff and by the IEG. After the project's approval by the Bank's board, the task team leader has to complete the project's first Implementation Status and Results Report[4] (ISR) within 12 months of the approval. After the first ISR, further ISRs are due at 12-month intervals until the project is closed (World Bank Independent Evaluation Group, 2009). Each ISR rates the likelihood that the project will meet its development objectives, and the Bank reports that projects with low ratings in mid-project supervision are more likely to have low ratings at closing (World Bank Independent Evaluation Group, 2009). Thus the ISR can provide an early indicator for project ratings (World Bank Independent Evaluation Group, 2009).

[3] QAG should be eliminated from the figure, as it does not exist anymore.
[4] In 2005, the Bank replaced the Project Supervision Report with the Implementation Status and Results (ISR) report. (Kilby, Assessing the contribution of donor agencies to aid effectiveness: The impact of World Bank preparation on project outcomes, 2012).

After a project is completed, the task team leader completes an "Implementation Completion Report (ICRs)". This is a self-evaluation of the effectiveness of the project in meeting its stated objectives. After 1995, all ICRs became subject to additional and independent validation by the IEG within three months after ICR is finalized, based on the evidence contained in the ICR. This process of secondary review by the IEG is called desk review, also known as "Evaluation Summaries" or "Evaluation Memoranda" (Denizer, Kaufmann, & Kraay, 2013). The ICR review assigns its own rating, which can differ from the Bank's self-ratings (World Bank Independent Evaluation Group, 2009). The IEG assessment tended to downgrade performance. The IEG rates, on average, 15 percentage points lower than the Bank's self-assessment in its final ISR (World Bank Independent Evaluation Group, 2009).

In addition to ICR desk reviews, the IEG nonrandomly selects 25–30% of the reviews and conducts in-depth reviews that involve field visits and interviews of multi stakeholders (World Bank Operations Policy and Country Services(OPCS), 2006). As a result, the IEG produces "Project Performance Assessment Reports (PPAR)". When the IEG selects projects for PPAR, preference is given to projects that are innovative, have potential for impact evaluation, have poor ICR quality or show disagreement with the ICR, those that are requested by Executive Directors or co-financers, or those that are relevant to upcoming studies or country evaluations (World Bank Operations Policy and Country Services(OPCS), 2006; Xu & Weller, 2009).

3. A quick look at some characteristics of successful and unsuccessful projects

To examine the factors that strongly contributed to the successful or unsuccessful outcomes of their projects, the most ideal way is to examine each project's implementation and completion report one by one and summarize the factors. Given the limited resources to conduct this study, I use a general approach using the quantitative data and some staff interviews. Nevertheless, to have an idea of what a successful or unsuccessful project would look like, I examined three projects that received highly satisfactory ratings, and three projects that received highly unsatisfactory and unsatisfactory ratings. Based on each of their ICR, I summarize a few common characteristics of satisfactory and unsatisfactory projects.

A primary education improvement project in Chile (Project ID: P006668) that was implemented between 1991 and 1998 was rated as highly satisfactory. The implementation completion report notes various reasons for its successful implementation. The project's aim was to enhance the quality of primary schools as well as preprimary schools and strengthen the institutional capacity of the Ministry of Education (MOE) in Chile. The project reached most of its targets, and some even exceeded the targets and eventually had a positive overall impact on the whole education system in the country, as evidenced by national standardized test scores. The borrower government showed high commitment to the project, the project was designed and implemented by the same team, which was good for maintaining the stability and consistency of the project. The establishment of democracy in Chile also created a favorable environment for the implementation of the project. The Bank also maintained a close relationship with the

borrower throughout all phases of the project and responded timely to any changes made in the project design as it was adapting to the needs of the local beneficiaries. As a result, the project served as a base for educational reform in Chile.

Other two Implementation Completion Reports of projects in India and Armenia that received highly satisfactory rating also indicated similar factors for project success. A technical education project implemented in India (P009988) was rated highly satisfactory as it received policy support from the government, as well as government commitment for the project. Field base project supervision with close interaction between the government and beneficiaries, and effective central coordination of multiple local states involved in the project contributed to the success of the project (World Bank, 2000b). Also, some factors are outside the control of the government and the Bank, such as rupee devaluation that led to additional funds and challenges in making timely adjustment of the budget affected project outcomes. An education financing and management project in Armenia (P008281) was considered highly satisfactory, mainly due to the continuous government commitment at the highest level, project designs that were limited to the immediate needs of the beneficiaries, stakeholder involvement in project design, the preparation and implementation phase, and the partnership between different ministries (World Bank, 2003a).

In sum, all three projects' ICR reported that government's commitment, particularly at the high level, and their ownership of the project led to the successful results of the project. Project design was also a factor that contributed to the success of the project. The government's leadership was also important when there were multiple implementing agencies and states involved. A project that has objectives that meets most

of the needs of the beneficiaries, stakeholder's involvement in the design and implementation mattered for the project's success. Consistency and stability of the implementation team, field based supervision, and outside factors such as changes in local currency or political form contributed to the outcomes of the projects.

The factors that mattered for projects that received unsatisfactory ratings were also similar to that of projects that received satisfactory ratings, however, external environment factors seem to have a significant influence on unsatisfactory projects. For example, a project that aims to improve manpower training as well as the quality and equity of the education system in Ethiopia (P000721), received unsatisfactory rating mainly due to two factors: complex design and weak implementation capacity. The "Christmas tree" design of the project, which led to less coherent project objectives disintegrated with the government's new policies, and not well defined objective component contributed to unsatisfactory results. Also, the government had a weak management capacity. The project ICR also notes that the circumstances were not favorable for the implementation of the project. The civil war intensified during the project period, and there was a change of personnel accompanied by the change of government, that led to serious delay in implementation. Although there was strong political support from the government even at the highest level, there were many challenges to overcome the delays.

A secondary education project implemented in Argentina (P064614) aimed to increase access in quality education and improve equity through school management received unsatisfactory results. The main reasons for unsatisfactory results included project design that had unverified assumptions, lack of a monitoring and evaluation

system and an incomplete assessment of risks. Despite the project implementation team having a strong capacity, some factors were outside the control of the borrower. The national bureaucratic procedure was slow, there was macroeconomic instability, and the new legislative framework was set in the education sector at the national level. At the provincial level, the law also changed, which led to weaker relevance of the project. Changing government priorities, changes in leadership at the sector level also created unfavorable conditions for implementation, resulting in the unsatisfactory rating.

A higher education project in Senegal (for the University of Dakar) received highly unsatisfactory rating (P002373). The project included basic provisions to limit student repetition, especially those who receive full scholarships, and professors' teaching obligations. Much of the budget was spent on student services in a context of high unemployment of university graduates. After a long national consultation process, the Government at the highest political level agreed at a package of reforms to address some of these challenges. The ICR explains many factors outside the government's control as reasons for project outcome. The president who was involved in making the agreement lost the election in 2000 and the new president, who made promises that he would drop the reform, did drop the reform, which may have had detrimental influence to the long-term development of the University of Dakar. Thus, the climate in universities was unfavorable to implement the project. The ICR describes that there was no political commitment and leadership for the project, and the government was institutionally vulnerable and the implementing agency was technically weak (World Bank, 2003b). Though this project was unsuccessful in terms of policy reforms, the project funded the construction of a modern university library that now is an example of other Francophone

African countries. Overall, the ICR report concludes that the borrower government's commitment, efficient implementation team, and project design that meet the local demand seemed to stand out for project's success. On the other hand, ICR report illustrated that factors outside the control of the borrower government seemed to influence more for projects that received unsatisfactory ratings (World Bank, 2003b). Even if the government showed political commitment for the project and implementation team had the appropriate capacity, external factors, combined with some factors within the borrower government or project design, led to unsatisfactory results. The factors listed from the ICR of these several projects suggest the factors that need to be included in my analysis.

4. The World Bank education policy over time (1960s–2011)

World Bank projects are the public face of its policies (Moloney, 2009). Reviewing the Bank's educational lending policy is important because it influences the characteristics of projects. The literature has long emphasized that policies are important factors influencing the implementation of the projects (Deininger, Squire, & Basu, 1998). Both external events and internal learning have influenced the project and policy agenda (Moloney, 2009). In this section, I review the World Bank's policy from the 1960s and examine how the general lending policy at the Bank has affected the education sector and the characteristics of education projects over time.

From the 1960s to the beginning of the 1980s: Manpower forecasting and technical and vocational education

The World Bank began to use human capital measurement for its analytic work in the 1960s (Heyneman, 2003). The Bank's original mandate of reconstruction for Europe began to shift toward newly emerging developing countries. The concept of infrastructure changed from bridges and highways to industry, agriculture and manufacturing. In order to maintain the industries and run manufacturing efficiently for long term, an investment in human capital was needed (Gavin & Rodrik, 1995; Heyneman, 2003).

From 1962 to 1980, the Bank, under McNamara, used the manpower forecasting models to justify all education investments. The manpower forecasting technique basically questions how much skilled manpower is needed to sustain an investment in industry, manufacturing and agriculture (Heyneman, 2003). The World Bank supported projects that provided skilled trainings since these were important aspects of education that improve production (Heyneman, 1985). Therefore, education lending was focused on vocational and technical specializations, while other areas were de-emphasized (Heyneman, 1999). Secondary education, in particular, was viewed as a road to "practical employment." The 1974 Education Sector Policy Paper suggests that education projects re-orient their content away from academic and toward vocational purposes (World Bank, 1974). As a result, between 1962 and 1980, 22% of all Bank education lending was devoted to 'diversified' secondary education (Heyneman, 1985).

However, a myriad problems arose in implementing "diversified" secondary education. These projects were 30–50% more expensive to construct and manage. There were "maintenance problems", including lack of materials and supplies as teaching

materials were categorized as “recurrent” costs that is considered as inappropriate for lending (Heyneman, 1985; Heyneman, 1999). There was also lack of teacher availability (Heyneman, 1985). Thus, utilization of the facilities was 58% on average (Heyneman, 1985). Such problems added complexity to implementation.

The education lending during this period also concentrated on capital improvements–especially school construction and equipping physical plants (Jones, World Bank financing of education: Lending, learning and development, 1992) (Mundy, 2002). Physical construction in the education sector took about 84% of available finance (Jones, World Bank financing of education: Lending, learning and development, 1992).

From 1981 to mid-1990s: Focus on primary education

Education lending focused on basic and academic education between 1981 and mid-1990s, which is the era of structural adjustment. As a consequence of the second oil shock in 1979, the growth of the world economy slowed and many developing countries faced a debt crisis (Tarp, 2010). The goal of poverty alleviation in the 1970s was replaced by structural adjustment (Jones, World Bank financing of education: Lending, learning and development, 1992), which aimed to downsize public expenditure, liberalization of markets and privatization of public utilities. Education lending, however, expanded between 1985–1994, as it showed better performance compared to projects in other sectors, and some governments proposed to allocate more resources to education sector (Mundy & Verger, Forthcoming). The 1980 Education Sector Policy Paper (World Bank, 1980) revealed that the Bank lending is not limited to technical, vocational education and diversified secondary education. Education quality was legitimate objective for lending,

and all levels of education, from elementary to higher education, were considered for lending (Heyneman, 2003).

However, during this period, there was a tendency that lending was focused on basic education. One of the rationales for the focus on basic education was based on the economic rate of return methodology (Heyneman, 2003; Mundy, 2002). Based on the rate of return analysis, the wage differentials by level of education showed that investing in primary education produces higher net returns on public investment than investing in secondary education, and secondary education produces higher returns than tertiary education. From this methodology, a "short education policy menu" emerged with its main message focusing on academic and basic education and shifting away from vocational and higher education. The short education policy menu also suggested a larger cost sharing of higher education through student loan schemes (Psacharopoulos, 1986; Moock & Harbison, 1988; Psacharopoulos, 1980; Psacharopoulos, 1990). Countries facing debt crisis were advised by the Bank to restructure their education sectors following this simple "short education policy menu" (Heyneman, 2003).

The 1990 Jomtien Conference on Education for All was another motivation for the World Bank to make a major shift toward primary education. Donor's neglect on primary education can be partly explained by their preference for supporting investment projects that are capital and foreign exchange intensive limited in scope so as to facilitate supervision and heavily dependent on donors' expertise in terms of technical assistance and training (Fredriksen, 1990, p.29). Since the World Conference on EFA in 1990, the World Bank has made effort to step up for EFA that includes sharp increase in lending in primary education and retraining of its staff. For example, the lending for primary

education in Sub-Saharan Africa increased from 29 percent of total lending as an average for 1985–89 to 46 percent as an average for 1995–1999 (Verspoor, 2001). The high investments in primary education definitely contributed to the shift in primary education, however, the promotion of universal primary education (UPE) and eradication of illiteracy were strong goals in themselves. Because of its strong priority for primary education and for achieving EFA goals, the World Bank's policies have often been considered to be a sign of reluctance to support higher education, particularly in Africa (Bollag, 2004). However, rather than being against higher education, the Bank's operational policy has generally been guided by the principle that support for education should be holistically integrated (Bollag, 2004).

During the latter half of the 1980s, the Bank increased focus on policy-based lending, which reflected a growing awareness on the support for policy and institutional reforms through lending instrument (Verspoor, 2001). There was also a pushback against the structural adjustment policies of international financial institutions in the late 1980s that affected the World Bank. By the early 1990s, there was a growing perception that policy conditionality was failing to reform policies and there was also a fear that aid can generate undesirable dependency in relationships (Killick, 1995; Mosley, Harrigan, & Toye, 1995; Kanbur, 2000; Svensson, 2003). Criticisms against multilateral financial institutions began to escalate and there was a growing rejection against structural adjustment (Xu & Weller, 2009; Jones, 1992; Mundy & Verger, Forthcoming). In an effort to respond to such criticisms, the Bank changed its organizational structure and procedures and increased the amount of policy-based lending (Heyneman, 2003). The Wapenhans Report in 1993 raised concerns about the underperformance of the Bank's

operations throughout the 1980s, and opened the door for NGOs to be involved in Bank's operations (Xu & Weller, 2009). The involvement of NGOs in Bank operations meant that the Bank had much more dialogue with international NGOs and access to national NGOs depended on the conditions of the country.

The criticisms also promoted the World Bank and academic researchers to start investigating the relationship between aid and economic growth. Within the Bank, there was also a growing concern that the Bank's education lending tends to rely on the economic rates of return methodology, and too much emphasis was placed on basic education (Heyneman, 1995; Heyneman, 2001; Mundy, 2002).

The components of education projects reflect the Bank's policy in the education sector. A majority of Bank lending projects implemented during this period focused on primary and academic education, with particularly successful projects in East Asian countries. New components, such as utilizing national assessments to increase school and teacher accountability, were added to educational operations by the mid-1990s (Laruch & Lockheed, 1992; Murphy, 1996; Benveniste, 1999; Mundy, 2002).

From the mid-1990s to 2008: The launch of Fast Track Initiative, complexity of education project portfolios, and the decline in education project ratings

The most prominent factor on World Bank's education lending during this period that is comparable to the 1990 Jomtien conference on EFA is the 2000 Dakar World Education Forum on Education for All. The 2000 Dakar forum proposed a Framework for Action to achieve the goals of EFA and MDGs by 2015. The forum had a major impact on the Bank's education sector, including a leading role in creating the Fast Track

Initiative (FTI, now the Global Partnership for Education: GPE), which was launched in 2002. The FTI is a global partnership between bilateral donors, international agencies and development banks to support universal completion of primary education (World Bank IEG, 2011). Participating countries benefit from better-coordinated aid and the funding was channeled through bilateral, multilateral as well as two FTI funds: the Catalytic Fund[5] and the Education Program Development Fund[6]. The World Bank hosted as a secretariat for FTI, supervising the Catalytic Fund and the Education Program Development Fund (World Bank IEG, 2011).

Between 1995 and 2005, the Bank also revitalized its focus on poverty alleviation. One of the prominent features of Wolfensohn's reform agenda was the inclusive participation of the borrower government in the design of national assistance strategies set by the Bank, thus giving more ownership to the borrowers[7] (Mundy, 2002; World Bank, 1999). Lending practices shifted toward a more client-centered direction that included a careful study of implementation of reforms (Mundy & Verger, Forthcoming).

The characteristics of education projects from 1996 and 2008 are more diversified than the simple "short education policy menu" during the structural adjustment period (Mundy, 2002). Within the context of global knowledge economy, the Bank called for greater public spending on higher education (Task Force on Higher Education, 2000; Samoff & Carrol, 2003). In addition, the Bank started to conduct experimental projects

[5] The Catalytic Fund provides transitional financial assistance to IDA countries who have education sector plans endorsed by donors but have difficulties mobilizing external funding at the country level because of relatively limited donor presence (World Bank IEG, 2011). The Fund provides three-year grants to support the scaling up of national education strategies. However, as of 2016, these do not exist anymore.

[6] The Education Program Development Fund supports technical assistance in the design of national strategies and capacities for countries to design a sound plan that can be supported by FTI.

[7] More ownership of borrower countries does not necessarily mean that the Bank put constraint in consulting borrower countries during the national strategy making process. Often the World Bank staff were not allowed to talk to citizens (e.g. teacher unions, NGOs, students, or journalists) without prior agreement from the borrower government.

on educational issues that were previously neglected (Mundy, 2002). The Bank also continued to play a major part in primary education, becoming the host of a multilateral pooled fund and a host for sectoral trust funds from bilateral and private sector donors (Mundy & Verger, Forthcoming). At the same time, the Bank focused on private sector investment in education, as there was increasing influx of private financing particularly among middle-income borrowers (Mundy, 2002; Mundy & Verger, Forthcoming).

With diversification of education lending, there is also a decline in education projects. The IEG (2011) report shows that education projects ratings have declined from 2000 and 2009. This contrasts with past performance, in which education projects always had better outcomes than other sector projects. The reason for such decline is an increasing number of projects in multiple sectors and has learning outcome agenda, and expanding support for post-primary education. The goals of these projects are difficult to achieve, compared to increasing access to basic education. Thus, the diversification and complexity of education projects and lack of its strategic sectoral focus may have led to the decline in project outcomes.

2008 – 2011: Financial crisis and continued emphasis on learning outcomes

One of the dynamics of the current World Bank is the rise of new major donors (Mundy & Verger, Forthcoming). As the economies of traditional donors faced financial crisis, addressing global poverty was not their top priority of policy agenda. Instead, China, India and Brazil, who are also the "clients" of the Bank, are now transitioning to "powerful principals" of the Bank, though they are mostly focused on expanding their geo political influence through bilateral aid (Mundy & Verger, Forthcoming). Another

dynamic is its shrinking group of IDA eligible borrowers. As some of the Bank's well performing IDA borrowers are set to "graduate" and become middle income countries, the Bank is left with a smaller number of IDA eligible borrowers (Mundy & Verger, Forthcoming). As the small group of IDA borrowers are mostly conflicted affected or politically fragile countries, it is possible that the recent decline in project outcomes can be due to countries' political/economic vulnerability.

The Bank released its new Education Strategy 2020 in 2011. It focused on "learning for all" through strengthening the education systems and building a high-quality knowledge base (World Bank IEG, 2011). The Bank plans to focus its effort on learning assessments, decentralization and promote involvement from the private sector. Although the new Education Strategy 2020 has been criticized for not revealing anything new (de Siqueira, 2012; Samoff, 2012), the strategy paper suggests future directions of the Bank's education sector policy. Moreover, since FTI was criticized for limiting focus on countries that showed good implementation performance and excluding countries (e.g. conflict-affected and fragile countries) that are in great need for aid, aid became more needs based.

5. Interim summary

This section consists of three parts. First, I explained how the scope of this study is narrowed from foreign aid in general to World Bank projects. Foreign aid started to help reconstruction of Europe in the aftermath of World War II and now assists the developing countries. Governments of developed countries, international organizations,

charitable foundations and religious philanthropies participate in providing foreign aid. Among these donors, only the World Bank has a direct input and output relationship data at project level, as it rated all projects implemented since the 1970s. The World Bank also has the largest funding source for education among all international organizations. This study narrows its scope to education lending, in which the source is from IBRD/IDA only. The study is also focused in examining the "black box," the relationship between project characteristics and project outputs.

The second part explained each phase of project cycle and possible issues that might arise at each stage. The project cycle is divided into preparation, implementation, and evaluation phases. The governments identify and prepare projects with support from the World Bank, and the Bank board approves. The preparation phase is assessed, in which the outcome is called "quality at entry." The local government implements projects and the Bank supervises and reports their progress. The budget is disbursed based on the estimates for annual disbursement in the Appraisal Report, which is modified as needed depending on implementation progress as established during supervision. When a project is completed, the Bank staff completes a report and the IEG reviews this report to rate the project. The rating is based on three criteria: 1) relevance 2) efficacy and 3) efficiency.

In the third part of this section, I reviewed the Bank policy since the 1960s. From the 1960s until the 1980s, the Bank focused on technical, vocational education, and emphasized practical secondary curricula based on manpower forecasting. During the period from 1980s to mid-1990s, educational lending was much affected by international movements such as EFA and rationale based on economic rate of return analysis, resulting in a focus on primary education sector and limitations in involvement in higher

education sector. From the mid-1990s to 2009, new donors emerged, project portfolio became more complex, and aid was focused on countries that showed commitments, becoming more needs based and performance based.

CHAPTER III

CRITICAL LITERATURE REVIEW

It is difficult to assess the "true" impact of aid on development, as researchers need an appropriate counterfactual that is difficult to establish without making assumptions, which are bound to be debatable (Tarp, 2010). Despite these challenges, many studies have examined the effectiveness of aid. In this section, I review the literature in three areas: 1) macro level studies on the effectiveness of foreign aid in general as well as in education sector 2) studies that used the IEG ratings data from economics literature 3) studies on success factors for project outcomes from project management literature, and 4) studies on World Bank education projects from education management literature.

1. Literature on aid effectiveness in general

There are two potential difficulties in measuring the effectiveness of aid. Reverse causality and counterfactual (Easterly, 2006). Reverse causality arises from the fact that countries receive aid because they are poor. Also, to properly measure the effect of aid, we need to be able to compare the treatment effect in two strictly independent situations – with and without aid (Tarp, 2010). This is the fundamental evaluation challenge, and in social sciences there is no way of addressing this problem without making assumptions that are bound to be debatable – in theory and in practice (Tarp, 2010). Despite the challenges, a great number of studies have been conducted on the effectiveness of foreign

aid (e.g. Cassen, 1994). I review some of the influential studies on effectiveness of foreign aid.

Cassen's (1994) book "Does Aid Work?" concluded that "much of the public discussion of aid has been distorted by prejudice, ideology and selective glimpses of parts of the evidence" and that "most aid does succeed in terms of its own objectives and obtains a reasonable rate of return; but a significant proportion does not." One of the early studies that showed no effect of aid on growth was in 1994, when Peter Boone, from the London School of Economics, wrote a working paper that found no significant effect of aid savings on growth. This showed that aid does not work on average. As one of the first to analyze the macro economic impact of aid in a neoclassical growth model, Boone found the evidence for negative rather than positive relationship between aid and growth (Boone, 1994). The study examined the effectiveness of foreign aid programs in relation to the political regime of recipient countries. Using data on nonmilitary foreign aid transfers, national accounts, human development indicators, and indexes of political liberties and political regime from 97 countries, this study used three alternative instruments, conducted instrument specification tests, and examined robustness using alternative subsamples and regression techniques. The findings of this study showed that in small countries, or countries where the aid/GNP ratio is extremely large (over 15% of GNP), aid does lead to higher investment. However, there was no significant correlation between aid and growth. There was also no significant impact of aid on improvements in infant mortality, primary schooling ratios nor life expectancy. The study found that aid increases government consumption. However, the study found that democratic/liberal political regimes have 30% lower infant mortality.

In 1997, the World Bank economists Craig Burnside and David Dollar (BD) published an important piece of study, a policy research report called "Assessing aid," which found that aid works in countries with good economic policies. The overview states "financial aid works in a good policy environment (and therefore) financial assistance must be targeted more effectively to low-income countries with sound economic management" (Burnside & Dollar, 1997). Their cross-country regression shows that foreign aid had no impact on growth in countries with poor macro-economic policies, while it led to faster growth in countries with good policies (Dalgaard & Hansen, 2001). Although the findings of this work have been controversial, this study has been the most influential study in shaping both the policy debate and research focus on aid effectiveness since Cassen's (1994) work (Tarp, 2010).

Many researchers tested the findings of Burnside and Dollar's study (Collier & Dollar, 2004; Collier & Dehn, 2001; Guillaumont & Chauvet, 2001; Hansen & Tarp, 2000; Lensink & White, 2001; Dalgaard & Hansen, 2001). They used the same data but applied different methodologies, such as including additional control variables and using non-linear specifications. Some of the findings corroborated Burnside and Dollar's results, but others found that the BD's results were fragile to changes in the sample. One of the researchers who challenged Burnside and Dollar's finding were Dalgaard and Hansen, who argued that the policy selectivity were very fragile, being extremely data dependent (Dalgaard & Hansen, 2001). They used the same data set as Burnside and Dollar, and found that small changes in the sample changed the result of Burnside and Dollar's findings. Therefore, they conclude that Burnside and Dollar's hypotheses were not robust.

Easterly, Levine and Roodman's (ELR) working paper for Center for Global Development (2004) also challenged Burnside and Dollar's work. They reconstructed the Burnside and Dollars data from original source and added additional countries and observations to this data set, and also extended the data through 1997 (BD's data end in 1993). Whereas other studies that retested BD's results changed the model specifications, their study maintained BD's methodology, but only changed the data set a little bit. Despite small changes in the data set, their results showed that Burnside and Dollar's results were fragile, thus supporting Dalgaard and Hansen's (2001) argument.

In response to these critics, particularly Easterly et al.'s (2004) study, Burnside and Dollar (2004) used similar data, which Easterly et al. used, and conducted the analysis again. With their own finding, they concluded that Easterly et al.'s (2004) argument is too negative to draw, and argued that successful evidence show that good policies contribute to faster growth. In addition, Collier and Dollar (2004) found that aid works best in countries with stronger government institutions and economic policies. An earlier study by Svensson (1999) argued that foreign aid raises growth in democratic countries.

As stated by the paper by Clements, Radelet and Bhavnani (2004), "Counting chickens when they hatch" can be called the first of a new generation of studies (Roodman, 2007). Their paper is different from other studies in that they narrowed the aid variable to parallel the economic growth. They showed that if they decompose aid variable and leave out humanitarian assistance and "long impact" assistance such as education and health, they can find a positive impact on growth. Their "short impact" aid variable includes budget and balance of payments support and aid for infrastructure and

industry, which can reasonably be expected to affect economic growth within a few years (Ranis, 2010).

Whereas Clemens and his colleagues (2004) were concerned about the mismatch between the aid and growth variables, the IMF economists, Rajan and Subramanian (2005), were concerned about the techniques used to remove reverse causation and other forms of "endogeneity" of aid variable, which is choice of instruments. Aid is an endogenous variable because poor countries can receive more aid, thus foreign aid increases when the GDP per capita decreases. This leads to a downward bias of foreign aid effect on the economic growth variable; therefore, it is possible that the actual effect of foreign aid can be positive and significant. Rajan and Subramanian (2005) saw that past papers used instruments such as policy interacted with aid; however, they argued these instruments do not properly correct for the bias from endogeneity. Rajan and Subramanian created a new kind of instrumental variable and with this, they found that there was no reliable effect of aid on growth. They concluded that aid not only had a negligible effect but can also have a negative impact on development.

On the other hand, Bruckner's (2009) study found a positive effect even after having addressed the reverse causality issue of foreign aid. Using a panel of least developed countries over the period of 1960–2000, the instrumental variable estimates, which used rainfall as an instrumental variable to generate exogenous variation in per capita GDP of countries, showed that a 1 percentage point increase in per capita GDP reduced foreign aid by 3 percentage point on average. The finding also showed that after the downward bias of foreign aid is accounted for, foreign aid indeed had positive and statistically significant effect on per capita growth.

Despite the mixed evidence of aid on economic growth and development, there is a widespread agreement in the literature that aid has often been very successful at the micro-economic level (Tarp, 2010). The World Bank has done some rigorous project evaluations, and reports from these evaluations are generally encouraging (Tarp, 2010). Even the extreme critics, including Easterly, admit that micro financed projects have a positive effect. However, at the macro level, it is difficult to observe a positive effect of aid.

2. Macro level studies on the effectiveness of foreign aid in education

Empirical studies on general aid effectiveness do not clearly answer the question of whether aid is effective. Some studies found that aid works in a good policy environment (Burnside & Dollar, 2000; Collier & Dollar, 2004). However, other researchers found that aid has no effect or negative effect on economic growth (Boone, 1994; Dalgaard & Hansen, 2001; Easterly, Levine, & Roodman, 2004; Rajan & Subramanian, 2005). Despite a large body of literature on aid effectiveness since the 1980s, only a few studies focus on educational aid (Michaelowa & Weber, 2008; Michaelowa & Weber, 2006; Dreher, Nunnenkamp, & Thiele, 2008; Christensen, Homer, & Nielson, 2011). These studies have found positive effects of foreign aid on school enrollment and completion rates, although the magnitude of the effect was modest. As addressing the endogeneity of foreign aid has been the key issue of researchers who conducts study on foreign aid effectiveness, all of the studies that focused on educational outcomes also attempted to address the issue of endogeneity of foreign aid in education.

Michaelowa and Weber's (2006) study examined the impact of aid for education on primary school enrollment and completion rate, and found a positive but modest effect. They used a panel data from 1975 to 2000 in about 120 low and lower middle-income countries. The authors used the Generalized Method of Moments (GMM) method and the instrumental variable estimation method to deal with the autocorrelation of panel data and endogeneity of aid in education. The study found a moderate but positive effect of aid on school enrollment and completion rates. The effect was larger when coupled with indicators for good governance. However, the positive effects were not robust in other model specifications and the authors were also uncertain about the validity of instruments (lagged education aid and energy aid).

Dreher et al. (2008) conducted a study similar to that of Michaelowa and Weber (2006), and found a robust positive effect of aid on school enrollment. They used panel data from 100 countries from 1970 to 2004 and looked at five-year average net school enrollment as their dependent variable. They used aid per capita (commitment[8]) as their explanatory variable and conducted pooled time series regressions controlling for lagged outcome and including country fixed effects. Like Michaelowa and Weber (2006), they applied GMM estimator and instrumental variable method, although Dreher et al. (2008) used different instruments.[9] Dreher et al.'s (2008) results showed a robust positive effect

[8] There are two different kinds of aid variables measured by the OECD-Creditor Reporting System. Whereas commitment is the amount of foreign aid that donors promise to contribute, disbursement is the amount that is actually allocated. Although disbursement seems to be a better measure, Dreher et al. (2008) use commitment for their main analyses because it has less missing data. They also explain that commitment is not that different from disbursement. Thus, they use disbursement data to check for the robustness of the aid effect and not for the main analysis.

[9] They used the international country risk guide, the index of economic freedom (governance related variables) and child mortality rate as instruments. Whereas Michaelowa and Weber (2006) explained that any health related variables (e.g., infant mortality rate) were correlated both with aid and enrollment, Dreher and his colleagues explained that mortality rate is uncorrelated with enrollment once other covariates are included in the model.

of aid on school outcomes. They also found that aid and its interaction with democracy was not significant, which is a contrasting finding to Michaelowa and Weber's (2006) study. However, they acknowledge that their 2SLS regression results may be biased as they failed to find a strong instrument.

Michaelowa and Weber's (2008) study used the same panel data and same method from their 2006 study, but looked at the impact of aid on all primary, secondary and tertiary level enrollment and completion rates. Their results generally showed that there were some positive effects of aid at all three levels. At the primary and secondary level, there was some evidence of positive effects, however, the magnitude of the highest significant positive coefficient was rather small. The effect of aid on tertiary gross enrollment was insignificant for the short-term panel, but there was some evidence of an increase in enrollment in long term. These results, nevertheless, were sensitive to different model specifications. Michaelowa and Weber (2008) concluded that these problems could not be solved econometrically and explained that the data problems led them to underestimate the effects of aid.

Michaelowa and Weber's (2006; 2008) and Dreher et al.'s (2008) studies showed that good quality data is essential to measure the impact of aid on education. Many missing cases in completion and enrollment rate remain an issue for their analysis. Since aid data disaggregated to sector allocation is only available from 1990s, Michaelowa and Weber's (2006, 2008) and Dreher et al.'s (2008) analyses were limited to assessing the impact of aid on *all* levels of education, not specifically at the primary, secondary or tertiary level.

Christensen et al. (2011) used a different data set (AidData 1.9) and applied a different methodology from previous studies. Christensen et al. (2011) separated out the effects of primary school aid from total aid in education and examined the effects on primary enrollment. They also disaggregated the aid data into two delivery channels: bilateral and multilateral. Christensen et al. (2011) conducted an Hierarchical Linear Modeling (HLM) analysis, which accounted for unobserved developmental factors that were clustered by country. Since HLM does not address the endogeneity concerns that arise from reverse causality, they estimated a separate model to find the predictors for aid allocation, and controlled for those predictors (democracy and corruption index) in the estimation of the effect of primary aid on enrollment. The results from Christensen et al.'s (2011) study showed that bilateral aid flows had positive and significant effects on enrollment rates, whereas multilateral aid flows were insignificant. Their findings were robust to the GMM model used by Dreher et al. (2008), but they also failed to find adequately exogenous instruments.

A more recent study in examining the effect of educational aid on school enrolment and educational quality is conducted by Birchler and Michaelowa (2013). Their study was conducted using the five-year panel data on school enrolments from 1996 to 2010 with wider data coverage available. The authors used the same panel data as in Michaelowa and Weber's (2006) study with same control variables but replaced the aid commitment data with aid disbursement data, which should be a more accurate measure. They also used the disaggregated data on education aid. The study found significant and positive but moderate effect of education aid on increasing primary enrolments even with country and period fixed effects. They also found positive effects

of education aid disaggregated by subsectors but only aid in education facilities and training remained robust to model specifications. The study also found a complementary effect of primary and secondary education. By using a qualitative comparative analysis, the study's finding also suggests a need for balance between primary and higher levels of education. The authors, however, acknowledged that the issue of reverse causality of foreign aid in education remained in this study since the time period with sector specific disbursement data was still very short.

Although the research on aid effectiveness and education aid effectiveness focuses on addressing the endogeneity of aid, another issue is the indirect relationship between education aid and educational outcomes particularly if they are measured at the country level. As foreign aid is delivered in the form of projects, researchers need to observe aid effect at project level, where the input and output relationship is more direct than country level studies. Moreover, as the macro level studies show rather inconclusive results, it can be more practical to assume that foreign aid works in certain circumstances and then explore *what* works rather than studying *whether* it works. Based on the assumption that aid works with certain project characteristics or specific circumstances, we should investigate the factors that make education projects successful. The next section discusses studies that used the World Bank project performance rating data, but the projects in the studies are not limited to the education sector.

3. Studies that use the World Bank project ratings data

Country level studies

Many studies focused on the effect of macro-economic factors on project performance (e.g. Isham, Kaufmann, & Pritchett, 1997; Isham & Kaufmann, 1999). Isham et al. (1997) ran a probit model using the project performance data from the Bank, and found that there was a strong link between civil liberties and the performance of government projects. These findings were robust using a variety of controls including measures of democracy. Isham and Kaufmann (1999), through Tobit analysis of economic rates of return, found that with good policies the productivity of investments increases. Policy indices used in the analysis include black market premium, fiscal deficit and index of trade restrictiveness. Other variables such as GDP growth, institutional indicators, and project level variables, such as project complexity and the size of public investment program, were included in their model. These studies, however, did not examine detail characteristics of projects.

Dollar and Svensson's (2000) study showed that political-economic factors are important for structural adjustment program outcomes. They focused on structural adjustment programs during 1980–1995 and used the IEG data, which had 220 structural adjustment programs. They applied a linear probability, probit, and two-stage probit models. Using a variety of macro-economic data from different sources, they found that political economic factors influence the success or failure of adjustment loan programs. The implications of this study, however, are limited to structural adjustment loan programs at the Bank.

As corruption is a serious issue in delivering aid, the most important political conditions for aid to work well are good policies and good institutional environments (Burnside & Dollar, 2000; Collier & Dollar, 2004). Dollar and Levin's (2005) study shows that high quality institutions in a recipient country increase the probability that aid will be used effectively. Dollar and Levin used World Bank project data from 75 countries between 1990 and 1999. Since they were conducting a country level analysis, their dependent variable was the proportion of World Bank projects that were determined to have "satisfactory ratings" in each country. The institutional quality was measured using the Rule of Law index, Freedom House democracy measures and a single overall index of institutional quality. They applied an Ordinary Least Squares (OLS) and instrumental variable estimation[10] because they wanted to address the issue of reverse causality (e.g., successful projects can build strong institutions). Their study showed that there was a strong relationship between institutional quality and project success rate. The study also showed that the regional dummy for Sub-Saharan Africa was a robust indicator for low success rate.

Chauvet, Collier and Duponchel's (2010) study found that being a post-conflict country when the project starts increases the probability of project success, but the positive impact gradually fades. Using the IEG data, they used probit and ordered probit models to estimate the probability that Bank projects would be successful, depending on a set of project and country characteristics, and the factors relating to the history of civil war countries. Their project characteristics included the duration of the project, IEG

[10] They instrumented the institutional quality using a variety of instruments that were used in other studies including the share of population that speaks English, the share of population that speaks a continental European language, the distance from the equator, and the size of population. The Sargan test statistics measuring the weakness of instruments were 2.7–2.8.

evaluation on the quality of preparation and supervision, type of loan, and source of Bank funding (IBRD/IDA). Regional and decade dummies were also included in the model. In addition to the positive impact of being a post-conflict country on project outcomes, they also found that project preparation and supervision is crucial elements for the success of projects, yet the impact of supervision on project outcomes is larger than that of preparation. Their results from sector analysis shows that education sector projects, which usually have a higher rating than other sector projects, tend to be less successful in post-conflict situations.

Guillamont and Laajaj's (2006) study shows that a project is unlikely to succeed in an economically vulnerable country because of the unstable environment, but when the ODA reaches 16% of the Gross National Income (GNI), the negative effect is compensated. They created an index of instability and their regression model controlled for country variables, project variables (sector of the project, a dummy for IDA/IBRD, a dummy for investment/adjustment loans) and added year fixed effects. Based on their findings, they argue that in an economically vulnerable country, the success of the projects increases with the level of ODA because of the cushioning effect of aid.

Dreher et al.'s (2010) study shows that the economic vulnerability, among other factors such as weak institutions and internal conflicts, tend to have lower project success as a consequence of political influence. Their study aimed to test the general hypothesis that aid given with political intention is not effective. Using the Bank project ratings data, they conducted conditional logit models with year and country fixed effects. They found that, on average, the projects that are politically motivated, such as those granted to governments holding a non-permanent seat on the United Nations Security Council or an

Executive Directorship at the World Bank, do not necessarily receive lower ratings than other projects at the Bank. However, when countries are economically vulnerable in the first place, projects were likely to get negative outcomes.

The studies that examined country level factors did not include detailed characteristics of the projects, and are not focused specifically on the education sector. However, these studies indicate a good reference for country level variables to include in my analysis.

Project level studies

There are relatively a small number of studies on the determinants of project characteristics of World Bank project outcomes. Denizer, Kaufmann and Kraay's (2013) study is closely related to my study. They investigated country level and project level correlates on project outcomes, using the IEG data on project ratings between 1983 and 2011. Their project level characteristics include detailed features of projects unlike the other previous studies. These include whether the project spans multiple sectors, the extent to which the project is novel, preparation and supervision costs, early warning indicators, and the quality of task team leaders who were in charge of each project.

Their findings from OLS estimations show that among country level variables, measure of policy and institutional quality were significant, which supports the previous literature. Interestingly, the data show that the project outcomes varied much more within countries than among countries. Among project level variables, they found that project preparation, supervision, early warning indicators, and task team leader characteristics are significantly correlated with project outcomes.

The authors were, however, concerned about the unobserved project level factors that might be correlated with project outcomes (e.g. problematic projects require greater supervision, and are more likely to get lower ratings). Since it is unlikely to find an appropriate instrumental variable, they quantified the magnitude of the likely biases in their OLS estimates, using Bayesian methods. Their results show that there was only modest magnitude of possible bias, which was not sufficiently large to influence the outcomes. The results suggested that improving project preparation and supervision can change the project outcomes. However, their study examines project ratings across all sectors, and do not specifically focus on education.

Kilby's (2012) study specifically examines the impact of preparation time on project outcomes. The IEG data was used, from which he samples 4,147 projects from 1989 to 2011. To generate the measure of project preparation time, the author used stochastic frontier model with political interest variables such as UN voting alignment, UNSC non-permanent membership and World Bank Executive Board membership, which are likely to accelerate the preparation time, but are exogenous to project quality. Conditional logit estimates with country fixed effects are used for project performance. The study found that projects with longer preparation time are significantly more likely to have satisfactory ratings, and the impact of project preparation time was greater in countries with vulnerable economic conditions. This study overcame the issue of endogeneity of project preparation, by taking advantage of political influence on preparation process that are not correlated to project outcomes. However, again, the study does not specifically focus on education, and does not include many project characteristics.

Moloney's (2009) study focuses on the World Bank public sector management projects. The study investigates whether the World Bank's public sector management projects improve governance. The author is interested in the project cost and whether having more public sector management projects within a country improves governance. Using the World Bank project database for explanatory variables and World Bank governance indicators for dependent variables as a measure for governance, the author employed panel corrected regression models using random effects. The general finding suggested that public sector management projects did not have a positive or negative impact on governance. The author explained that such finding could be because public sector projects are influenced by politics as well as data incongruities. Moloney's (2009) study is one of the very few studies that examine aid effectiveness at the project level; however, it focuses on public sector management projects. The outcome of interest in this study is not on project ratings but on governance indicators, including government effectiveness.

Hassan's (2012) study observes the effects of external environments such as government and economy on World Bank project ratings across all sectors. Using the IEG data and country level data from various sources, the author conducted principle component analysis to identify components and used logistic regression to find a relationship between the components and project outcomes. Results show that good governance and a high level of industrialization were strongly related to project performance. The author also conducted regression analysis by sector and found that governance was most important for education sector projects. The study also shows that World Bank supervision is highly associated with project outcomes. However, the study

uses Quality of Bank Supervision ratings[11] to measure Bank supervision, which may not be an accurate measure since the outcome ratings may be influenced by sub-ratings including Quality of Bank Supervision. Thus, further analysis is needed to conclude that Bank supervision can offset influences from external environments on project outcomes. Although this study includes sector analysis and shows some findings from education sector, the study only focused on external factors and did not include project-specific factors (i.e. project cost, loan type, subsectors) in the analysis.

Studies on political influence

There are some studies that examined political influence on allocation of funds at the World Bank, and on project characteristics (Frey & Schneider, 1986; Fleck & Kilby, 2005; Kaja & Werker, 2010; Dreher, Klasen, Vreeland, & Werker, 2010; Kilby, 2013). Studies found that the World Bank shareholders' political interests significantly influenced the Bank's allocation of funds, including geographical distribution and IDA lending (Frey & Schneider, 1986; Fleck & Kilby, 2005; Andersen, Hansen, & Markussen, 2006). However, other studies that focus on IDA lending found that there was no significant association between donor interest variables or board membership and

[11] Kilby (2014) points out some issues with the endogeneity of using sub-ratings in the IEG data such as "quality of preparation" and "quality of supervision." Kilby (2000) and Isham et al. (1995) discuss the "halo effect, " which is a common problem with the World Bank project performance data. "Halo effect" occurs when the measurement of one variable is affected by the observed state of another variable (Isham et al., 1995). Outstanding performance in one dimension or characteristic may tend to bias upward the evaluation of other dimensions or characteristics (Isham et al., 1995: p. 187). The World Bank first introduced overall ratings, and then later, they introduced process ratings. If an evaluator sees a project whose overall rating is unsuccessful, it is likely that an evaluator will rate quality of preparation and supervision and other performances unsuccessful based on the first impression of overall quality of the project. Thus, there is a project-specific evaluator bias common to the overall ratings and sub-ratings. If one evaluator did the overall rating and a different evaluator did the sub-ratings, the correlation between these measurement errors will be eliminated and the problem would be solved (Kilby, 2014-email communication). However, the problem remains when the evaluator for overall ratings and sub-ratings are the same. From my preliminary interview with the Bank staff, I regard the concerns for "halo effect" is not negligible, therefore I do not include the sub-ratings as explanatory variables in my study.

IDA loans and grant commitments (e.g. Kaja & Werker, 2010). IBRD loans were influenced by political interests, as countries received a large increase in loans when they had a seat on the board (Kaja & Werker, 2010). However, these studies focus on the relationship between political interests and project characteristics (size of loans, IBRD/IDA), but do not examine its relationship with project outcomes.

At the project level, there are two studies that examined political influence on the number of projects a country received and on the preparation time. Being a UN Security Council (UNSC) member increased the probability of receiving World Bank projects, although it did not affect the size of the loans (Dreher, Sturm, & Vreeland, 2009). Kilby's (2013) study investigates whether donors' political interests reduce the length of time from project identification to approval. Donors' political interests were measured combining a variety of information such as the recipient country's UNSC status, UN voting alignment, World Bank Executive Board membership, trade flows, military aid and bilateral aid. Evidence suggested that project preparation is accelerated for countries that hold geographically important position as a seat on the UNSC, and for countries that have a seat at the World Bank Executive Board.

Project outcome studies focusing on education sector

Very few studies specifically focuses on education projects. Kaufmann and Wang's (1995) study is the closest study as it focuses on social sector projects. Their study investigates the effect of macroeconomic policies on the probability of project failures in social sectors, which consist of education and health sectors. They use the IEG data, which has 259 social sector projects, and also use an aggregate data reflecting

economy wide policies. The project data covers 84 countries between 1974 and 1990. A variety of indicators are used for economy wide policy indicators, including currency over evaluation measured by black market premium, an index on trade restrictions, and fiscal deficit as a share of GDP. The results from the multivariate probit estimations presented that the indicators for macroeconomic instability were significantly related to the probability of project failure, and the impact was larger for education projects. The analysis also showed that project sustainability was largely affected by the macroeconomic policy indicators. Their study suggests the importance for economic wide policies for social sector projects; however, the study did not include variables that represent project level characteristics.

Vawda et al.'s (2003) study is the only study that focuses specifically on education projects, though it examines the relationship between the quality of economic analysis and project performance. They hypothesized that if a project's economic analysis is good or excellent, it will be easier to implement and will tend to receive a higher rating on the outcomes. They used a sample of 104 education projects between 1993 and 1998, which have been evaluated based on economic analysis. Their multivariate logistic regression analysis showed that the quality of economic analysis is significantly associated with project outcomes controlling for region. The implications of this study are, however, limited to short term results of project outcomes as their outcome data does not incorporate post-complete evaluations. The study also does not include characteristics of education projects.

The World Bank IEG report (2011), which examined the education project performance and its portfolios during 2001–2009, suggests many important variables to

be considered for my analysis. The report is based on evidence from the Bank's internal database that tracks project portfolios, desk reviews of project completion reports and evaluation documents for all projects managed by the education sector and by other sectors that have large education components. The report also used coded data of the statement of objectives for each education projects and also used the findings from past IEG evaluations.

Their descriptive findings from the data mainly illustrate that the education sector projects ratings have declined. The share of satisfactory performance declined from 82% to 69% between 2001 and 2009. The report shows that the cause of this decline is increasing number of projects in multiple sectors, expanding support to post-primary education in low-income countries, and the learning outcome agenda that increase the complexity of portfolio. The changing components of the education project portfolio provide challenges in evaluating projects. For example, the objectives of improving learning outcomes are difficult to achieve compared to the projects that aim to increase access to schools. Expansion of post-primary education is also not easy to implement quickly. In addition, the report notes that project design and preparation continue to play a key role in project performance. The degree of complexity of project design in relation to borrower's capacity, risk assessment and mitigation plans, level of political commitment, and the realism of time frame for results are all important components for project outcomes.

4. Project management studies

In the third part of the literature review section, I review the studies that examined the success/failure factors for project outcomes from the project management literature. Examining the project determinants from the international development project management literature will give a broader perspective in observing important determinants for outcomes of projects.

Studies pointed out the importance of project design, management and supervision for successful project outcomes (Ika, Diallo, & Thuillier, 2010; Ika, Diallo, & Thuillier, 2012). Ika et al. (2010) analyzed the extent to which the project managers use tools, techniques and available methods for project implementation and monitoring is associated with project success. Using the data collected through questionnaires from 600 respondents from 26 countries in Africa, they conducted factor analysis to determine the concepts, tools, and techniques of international development projects. They conducted a correlational analysis between project success and project managers' efforts to use tools and techniques. The findings showed that project design and characteristics combined with well monitored and evaluation tools have a higher chance to succeed. However, the analysis is based project outcome that is self-reported, thus the study is not free from self-report and non-response bias.

Ika, Diallo and Thuillier's (2012) study investigated the success factors of World Bank projects. They collected data from survey questionnaires from World Bank project supervisors and task team leaders who randomly chose a single project; thus the sample was 178 projects. They conducted exploratory factor analysis to identify critical success

factors and ran regression of project success on these factors. Their findings showed that monitoring, coordination, design, training, and institutional environment account for 64% of variance. They found that among these, project design and monitoring are particularly important for project success. On the other hand, they found there is a lack of statistically significant relationship between project success and characteristics of projects (e.g. size, duration, budget, etc.) and project supervisors. Interestingly, project success tended to be rated lower in education sector. The study, however, is also subject to self-report measures, hence there could be a possible bias that arises from the convenience sampling method and a bias resulting from non-respondents.

Diallo and Thuillier (2005) and Khang and Moe's (2008) studies imply that trust, communication and consultations among the staff (officials at donor agencies, coordinators and task team members) and stakeholders are significant determinants of project success. Diallo and Thuillier's (2005)study examined the success factors of international development projects in Africa, specifically focusing on trust and communication between coordinator, task team and stakeholders. The study is based on the data collected through questionnaires completed by 93 project coordinators of development projects in Africa. They operationalized the construct of trust and communication by identifying latent variables that drive interpersonal relationships between stakeholders and within the project team, through factor analysis. Findings from multinomial logistic regressions illustrate that communication and trust between the project coordinator and task manager relationship is the key factor for project success, whereas team cohesion is the second most important factor. Although their findings emphasized the need for a high quality of communication and trust between the

coordinator and project implementation team, a limitation of this study is that the project outcome is self-evaluated by respondents. Moreover, the interpersonal relationship among stakeholders, coordinators and task team members are limited to those perceived by project coordinators.

Khang and Moe's (2008)study analyzes success factors for international development projects at each phase of project cycle. Using empirical data from a field survey in Vietnam and Myanmar, the authors propose a conceptual framework of a project cycle, in which they defined key activities, key players and end products of each phase of the project cycle. The study aims to empirically validate this conceptual framework with a survey response collected from project managers, staff members and officials at donor agencies, government agencies and international non-governmental organizations. With 368 returned questionnaires, they analyzed the relationship between the success factors and project outcomes at each phase of project cycle and also compared this to overall project outcome. Their regression results shows that effective consultation with stakeholders proves to be the most influential factor on project management success and more important than the competency of project supervisors and managers. However, like Diallo and Thuillier's (2005)study, the definition of project "success" in this study is rather subjective, as it is based on the perceived judgment of survey respondents.

The studies from project management literature indicate the importance of examining project design, monitoring, trust and communication within the project team and between team and the stakeholders. As I continuously pointed out, these studies use project outcomes and variables that are self-reported, therefore, are subject to self-report

bias. Also, no study specifically focuses on success factors for educational development projects. Next, I move on to the education management literature and review studies that describe factors for success/failure of World Bank education projects.

5. Studies on World Bank education projects

The majority of education literature focus either on individual school level or education policy, and very few studies focus the issues at the project or program level, particularly in the context of educational development. In this section, I review a small number of studies in international educational development and educational management literature that investigate or evaluate World Bank education projects. Although managerial factors or project characteristics are not directly examined in these studies, the studies provide basis for understanding the mechanism and complexity of planning and implementing education projects in context of developing countries.

Many studies indicated that time and cost overruns are the issues in implementing education projects (London, 1993; Khan, 1995; Jain, 1997; Wang & Bergquist, 2003). Cost overruns occur due to macro-economic changes in domestic economy. London (1993) conducted a case study of a failed project in Trinidad and Tobago. Data was gathered from government files and progress reports of the project, World Bank documents, and commentaries on the theory and practice of education planning. The study described that the Short Term School Building Project was implemented when society was undergoing some significant changes due to influx of "easy petroleum dollars" into the local economy. This influenced the prices of land for

school sites, labor costs, and project management. However, the schools were still constructed with severely increased costs that led to huge cost overruns (London, 1993). Thus the study concluded that failing to redesign the project led to huge cost overruns.

Khan's (1995) study examined the education and training project implemented in Barbados between 1986 and 1990. Khan describes that there were time overruns due to some lack in the planning and design phases or some unexpected events. Time overruns occurred because the project needed to be redesigned to reduce costs and this delayed implementation of projects. Also, there was a heavily rainfall during the year of constructing school building and this led to unexpected delay in implementing the project.

Whereas other studies examine challenges of cost overrun, Jain (1997) analyzed success factors from two case studies of primary programs implemented in Bangladesh. The data was gathered through school visits and interviews with staff. Finances and cost control were among the eight elements the author listed for success. The author describes that significant resources were allocated to monitoring. Also, the donor commitments were stable over years of project implementation. Even though the author's definition of 'success' is based on student attendance and not on student learning outcomes, the study indicates the need to consider allocating sufficient resources for monitoring and ensuring financial stability from donor organizations.

Wang and Bergquist (2003) points out another issue of project funding. They examined three basic education projects implemented in China during the 1980s and 1990s. They raised the concern that project funding that was supposed to be burdened by the central or provincial government was being passed on to local communities and

households. However, they did not have the direct data to prove this; however, they present the ratio of financial allocations of state, province and local self raised funds for capital construction investment in education.

In addition to project funding issues, studies also point out the significance of project team and personnel issues (Khan, 1995; Jain, 1997). Khan's (1995) study examining projects in Barbados discussed the lack of the project team's internal consistency and reliability in monitoring and control of the project. It led to a failure in redesigning the project, unscheduled delays and lack of technical supervision. Jain's (1997) study, which examined the success factors of projects in Bangladesh, indicated that more authority was given to the operational personnel than headquarter staff. The two studies imply that more decision making by the operational staff and consistency within the project team in controlling and monitoring projects are crucial for successful outcomes.

The studies in education management literature are based on case studies. Most of them evaluate studies descriptively using project documents. Whereas studies from project management literature generally pointed out that project design, monitoring and supervision were important factors; studies that examined World Bank education projects showed specific issues of implementation such as time and cost overruns. Literature from both project management and education presented the importance of collaboration within project team members.

6. Interim summary

I reviewed the literature in four different areas. First, I reviewed macro level studies that examined the effect of foreign aid in education on school enrollment and completion rates. The studies showed a positive but modest effect (Michaelowa & Weber, 2006; Michaelowa & Weber, 2008; Dreher, Nunnenkamp, & Thiele, 2008; Christensen, Homer, & Nielson, 2011). However, the endogeneity of foreign aid could not be solved in all of these studies, as they failed to find an instrument or validate their instruments.

Second, I reviewed studies that used the World Bank project ratings data. Many studies examined the effect of macro-economic variables on project outcomes (Isham, Kaufmann, & Pritchett, 1997; Dollar & Svensson, 2000; Chauvet, Collier, & Duponchel, 2010; Guillaumont & Laajaj, 2006; Dreher, Klasen, Vreeland, & Werker, 2010). Studies found that civil liberties, good policies and high quality institution in a recipient country increased the probability of successful project outcomes (Isham, Kaufmann, & Pritchett, 1997; Dollar & Svensson, 2000; Dollar & Levin, 2005). On the other hand, post-conflict situations, economic vulnerability, and macro-economic instability had a negative effect on project outcomes, particularly social sectors including education (Chauvet, Collier, & Duponchel, 2010; Guillaumont & Laajaj, 2006; Dreher, Klasen, Vreeland, & Werker, 2010; Kaufmann & Wang, 1995). Studies that focused on project level characteristics found that project preparation, supervision, early indicators and task team leader characteristics were significantly related to project outcomes (Denizer, Kaufmann, & Kraay, 2013; Kilby, 2012).

Third, I reviewed studies from international development project management literature. The studies examined success factors of projects in the field of international development. Project design and monitoring were particularly important for outcomes (Ika, Diallo, & Thuillier, 2010; Ika, Diallo, & Thuillier, 2012) as well as trust, communication and consultations among staff and stakeholders (Diallo & Thuillier, 2005; Khang & Moe, 2008). However, the studies were based on survey participants' perceived outcome of the project, thus they are subject to self-report bias.

Lastly, I reviewed studies from education management literature that discuss World Bank education projects. Studies pointed out the issues in time and cost overruns that led to delay in project implementation and project failure (Khan, 1995; London, 1993). Internal consistency of project team and decision-making at the operational personnel was also important element for success (Khan, 1995; Jain, 1997). However, case studies relied on project documents. Thus, there is no study that uses the project level data and focus on education project outcomes across countries.

CHAPTER IV.

CONCEPTUAL FRAMEWORK

1. Middleton's (1985) analytical framework on management of World Bank education projects

Middleton (1985) conducted a study which concentrated on the management of World Bank education projects. This study presents an analytical model that can be applied specifically for World Bank education projects. Middleton's (1985) analytical model is useful for my study as it helps to organize the complex set of factors and their inter-relationship and the patterns of the relationships of variables around the operation of education project management (1985). As shown in Figure 7, Middleton (1985) presents four categories of factors for management performance: project environment factors, task factors, management system factors, and operational management factors.

First, the project environment factors are the factors that affect the nature of project management. Middleton divides the factors into two theoretical concepts: social stability and policy stability. Social stability is the stableness of the society as a whole during the implementation of a project. Revolution, civil conflict or natural disasters are examples of instability. Policy stability is the stableness of educational policy and institutional structure during the project period. Social stability and policy stability are related, as sudden change of the government due to social instability such as revolution is likely to bring about changes in the structures and personnel within the government, which eventually will lead to delays in project implementation and bad project

performance. In addition to the social and educational policy stability stated by Middleton, I add other factors that may represent the educational landscape of the borrower country (i.e. public spending in education sector, school enrolment) in this study.

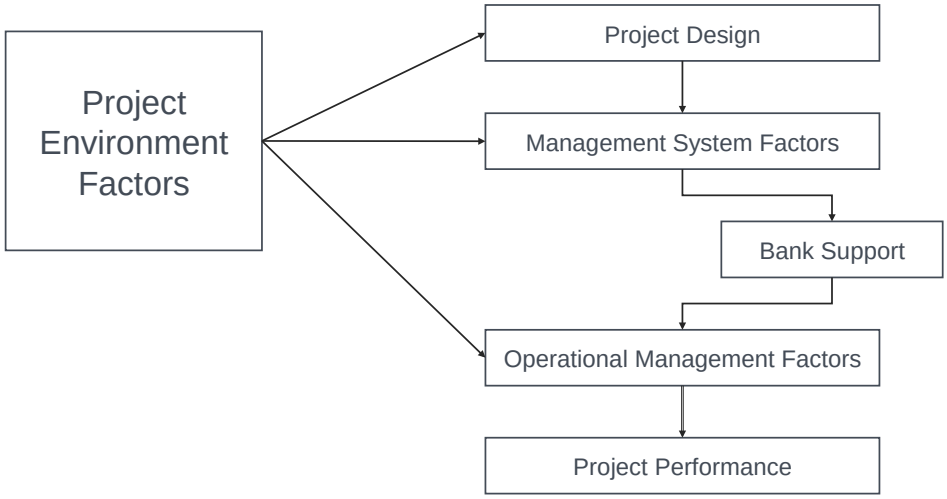


Figure 7 Concept map: categories of factors for project performance

External project environment factors are also mentioned in other empirical studies as they are significant factors for success of any international development projects. Table 1 shows that political situations of borrower country such as the degree of democracy (Dollar & Svensson, 2000; Dollar & Levin, 2005), post-conflict/post-civil war situations (Chauvet, Collier, & Duponchel, 2010), the degree of civil liberties (Isham, Kaufmann, & Pritchett, 1997) can be a determinant for project success. Economic stability or instability was also considered in several studies (Kaufmann & Wang, 1995; Guillaumont & Laajaj, 2006; Dreher, Klasen, Vreeland, & Werker, 2010; Kilby, 2012). As political and economic situations of a country can affect the social and policy stability,

I consider political, economic situations as well as social and policy stability as part of external project environment factors in my study.

Secondly, the task factors are referred as "project design variables." The goals/means of agreement are the extent to which the borrower government and the Bank agreed on the goals and strategies of a project. Although during the appraisal of a project the goals are agreed, they are negotiated goals, which means the policy differences between the borrower and the Bank remains, especially when there is turnover of borrower policy and management staff (Middleton, 1985). Thus, it can be hypothesized that the more goals are agreed during appraisal, the better a project performs. However, the goals and means of agreement cannot be measured in my study due to lack of data availability.

The second sub-element of task factors is "complexity and coordination." Complexity can be measured by components of sectors (i.e. academic, vocational), levels and institutions. Complexity of a project can be measured in my study using the information on project subsectors and percentage of the subsector for each project. Project design and complexity of projects are also noted by previous studies (Table 1). In general, project design was found to be significant for project success (Khan, 1995; Ika, Diallo, & Thuillier, 2012). More specifically, Ika et al. (Ika, Diallo, & Thuillier, 2012) found that the project size and duration mattered. Whether a project has a target that is set across multiple sectors and the novelty of projects were considered in Denizer et al.'s (2013) study. Preparation of projects was considered in multiple studies as well (Chauvet, Collier, & Duponchel, 2010; Denizer, Kaufmann, & Kraay, 2013; Kilby, 2012). "Innovation and demand," the third sub-element of task factor, is the degree to which a

project introduces new forms of education and the degree of social demand for the innovative services. It can be hypothesized that the more innovative a project is, the more difficulty a project faces with its management. Also if there is a lack of social demand for the innovative project, it can be a potential challenge to the project management (Middleton, 1985).

Table 1 Factors from the literature review categorized into Middleton's (1985) analytic model

Factors related to project manager and team (operational management)	**Factors related to the project design**
- Quality of task team leaders (Denizer et al., 2012) - Trust (Diallo and Thuillier, 2005) - Communication between project coordinator and task manager (Diallo and Thuillier, 2005) - Team cohesion (Diallo and Thuillier, 2005) - Consultation with stakeholders (Khang and Moe, 2008) - Project manager using tools and techniques available (Ika et al., 2010) - Coordination (Ika et al., 2012) - Training (Ika et al., 2012) - Internal consistency and reliability in monitoring and control of project (Khan, 1995)	- Size, duration (Ika et al., 2012) - Design (Khan, 1995; Ika et al., 2012) - Preparation (Chauvet et al., 2010; Denizer et al., 2012; Kilby, 2012) - Supervision (Chauvet et al., 2010; Denizer et al., 2012) - Monitoring (Ika et al., 2010; Ika et al., 2012) - Multiple sector (Denizer et al., 2012) - Early warning indicators (Denizer et al., 2012) - Novelty of project (Denizer et al., 2012) - Post primary education (World Bank, 2011)* - Learning outcome agenda (World Bank, 2011)* - Complexity of portfolio (World Bank, 2011)
Factors related to management system	**Factors related to external environment**
- Institutional environment (Ika et al., 2012) - More authority among the operational personnel (Jain, 1997) - Ensuring fiscal stability from donor organizations (Jain, 1997) - Corruption (King, 2007)	- Civil liberties (Isham et al., 1997) - Black market premium, fiscal deficit, index of trade restrictiveness (Isham and Kaufmann, 1999) - Measure of democracy (Dollar and Svensson, 2000; Dollar and Levin, 2005) - Post-conflict/post civil war situations (Chauvet et al., 2010) - Economic instability (Kaufmann and Wang, 1995; Guillamont and Laajaj, 2006) - Level of ODA (Guillamont and Laajaj, 2006) - Economic vulnerability (Dreher et al., 2010; Kilby, 2012) - Institutional quality (Dollar and Levin, 2005; Denizer et al., 2012) - Policy (Denizer et al., 2012)

The last sub-element of task factors is "institution and building needs." Middleton (1985) explains this as the degree to which the project is seeking to develop project management capacity or the institutional capacity of the project implementation agency, or both. The World Bank projects' general goal, in addition to the explicit goals, is institutional building of borrower countries. Therefore, institutional building is a common goal in many projects, even though they are not explicitly mentioned in official reports. Middleton (1985) hypothesized that if there is a higher need for institutional building; this indicates the institutional capacity is low, creating more difficulties with management tasks. In my study, the degree of innovation and social demand for innovative projects and the degree of need for institutional building cannot be obtained from the quantitative data; therefore, these factors can be captured in the qualitative interview data.

The next category is management system factors, which is related to any issues of structure of the management unit in the borrower country. Middleton (1985) lists four factors within this category: access to the top decision-makers in the management system, coordination and control by the project unit over the institutions in implementing the project, the number of staff and their skills, and the degree of continuity of key staff in the project units. High levels of these four factors should lead to effective management of projects. The task team leader quality was one of the factors that Denizer et al. (2013) also considered in their quantitative analysis; however, since there is no information on the four sub-elements of management system factors in my quantitative data, I explore these factors from qualitative data.

The last category of factors is operational management factors. The category consists of three sub-elements: input management, monitoring, and interaction with the

Bank. Input management is related to any issues regarding the delivery of supervision of project inputs such as civil works and procurement. Monitoring is issues associated with Bank monitoring and supervision as well as general monitoring information within the project. The third component, interaction with the Bank indicates the nature of the relationship between the borrower and the Bank during implementation. The effectiveness of communication and the degree of cooperation, which Middleton stated in describing the interaction with the Bank, is also emphasized in Ika et al. (2012) and Diallo and Thuillier's (2005) study. Monitoring and supervision was also an important factor in Chauvet et al. (2010), Denizer et al. (2013), Ika et al. (2010) and Ika et al.'s (2012) studies.

I apply Middleton's (1985) analytic model in my study even though there is lack of data availability to cover all the content areas mentioned in the theoretical model. The quantitative data (World Bank ratings and project characteristics data merged with country level indicators) have limitations in covering all four areas in the analytic model. Yet, the quantitative data should be able to cover some factors. The World Bank project database provides factors related to project design (subsectors and complexity of projects) and the World Development Indicators should be able to cover part of project environment factors, such as democracy and economic growth. Middleton's (1985) model can be applied more for the qualitative data analysis part of this study. In particular, a majority of factors mentioned in management system and operational management factors can be obtained from the staff interview data. Therefore, I use this analytic model in analyzing and coding the interview data and interpreting the inter-relationships between different factors in the discussion section.

Although Middleton's (1985) analytic model is an appropriate theoretical framework that fit my study, the author acknowledges some limitations of Middleton's (1985) study. The study is based on a small set of data that was available at that time (seventeen education project completion reports and staff interviews) and used the subjective project ratings, which can be influenced by the individuals who completed the reports. The data was chosen to represent a sample of education projects from all regions and attempted to select projects of different sizes and different levels of success (Middleton, 1985). This indicates that the sample was not randomly selected and implies a potential for sampling bias. Despite the limitations of Middleton's (1985) study, I reiterate that the model helps to understand the complicated process of World Bank education project implementation and management.

2. Interim summary

In this section, I explain a conceptual framework to examine determinants of World Bank education project outcomes using Middleton's (1985) analytic model. Middleton (1985) conducted a study on management of World Bank education projects using completion reports from a small sample of education projects and staff interviews. Figure 7 illustrates the different set of categories and components within each category suggested by Middleton. Project performance (or project management performance) can be influenced by four large categories of factors: project environment, task factors (project design), management system, and operational management factors. Project environment factors consist of social and educational policy stability that can be

influenced by political and economic situations of the borrower country. Task factors/project design variables are factors related to the project components. This category includes the degree of goal agreement between Bank and the borrower, complexity in terms of project components, the degree to which the project component represents innovative services and social demand for the innovation, and institutional building and needs including the Bank's support for the institutional capacity to develop project management.

The factors in the management system category are any issues related to the structure and function of the project implementing agencies, usually the borrower government. The background and skills of the Bank and local staff who are involved in project implementation and their continuity of being involved in the project operation matters for project success. Operational management factors include any input management (i.e. delivery of inputs, civil works, training, procurement), issues associated with monitoring of the project including Bank supervision, and interaction between the borrower and the Bank. The overall project management performance is hypothesized to be affected by these four categories of factors.

CHAPTER V.

RESEARCH DESIGN

1. Mixed methods design

In order to answer my research questions, I use a mixed method design. Mixed methods design uses both quantitative and qualitative data, and the integration of these data in a single study to answer a common research question (Teddlie & Tashakkori, 2009).

Using a mixed method design is particularly important for this study, since the quantitative data contains limited information on project characteristics and project performance outcomes are measured in ratings that can be weak to subjectivity of individuals who evaluated the project. In addition, by applying different types of methods I can neutralize or cancel the biases of other methods as all methods have limitations (Creswell, 2003). Using only a quantitative method can lead to a weakly contextualized study with limited policy relevance when policy makers are not only interested in whether an intervention (or a specific factor) worked but also in why and how it worked (White, 2008). Therefore, by triangulating data sources – a way to seek convergence across qualitative and quantitative methods (Creswell, 2003) – I will be able to draw a comprehensive picture of education project implementation.

In combining quantitative and qualitative approaches, I use "confirming/reinforcing," "enriching," "explaining" and "merging" defined by (Carvalho

& White, 1997). “Confirming/reinforcing” is verifying quantitative results through a qualitative approach (Carvalho & White, 1997). Since the staff interview was conducted after preliminary analysis of quantitative data, I was able to verify the initial results from quantitative analysis through staff interviews. “Enriching” is using qualitative data to obtain information on variables not obtained by quantitative data (Carvalho & White, 1997). “Enriching” is necessary in my study as the quantitative data has a limited number of variables on project characteristics and as there can be many factors that can be obtained only through qualitative data. “Explaining” is using qualitative work to explain any unexpected results from quantitative data. The staff interview data can help interpret and explain the results from the quantitative data. “Merging” is analyzing and integrating the information provided by quantitative as well as qualitative approach to draw one set of policy recommendations (Carvalho & White, 1997). Since my study aims to draw policy recommendations on educational development project management and implementation, using both methods can help suggest a consistent set of policy recommendations. While I attempt to “merge” the information from quantitative and qualitative data, it is difficult to draw many policy recommendations, as there is very small overlap of information from qualitative and quantitative data.

2. Research Question 1: Determinants for education project performance

Data

This study is based on a sample of 742 education projects approved between 1996 and 2011[12]. The data is obtained from the World Bank website. I merge the World Bank project performance data with the IEG project ratings data using the project ID. This data is then merged with country level indicators from various sources. I explain the details of each data below.

World Bank data

I use the World Bank project performance ratings as my data for outcome (the dependent variable) in this study. The World Bank project performance data has ratings from all project assessments carried out since the IEG was created in the 1970s (World Bank Independent Evaluation Group, 2012). The IEG assigns ratings using a standardized scale, and the ratings criteria have evolved through the years, ranging from a single outcome to several scales of outcome today (World Bank Independent Evaluation Group, 2012). The World Bank performance ratings data have several advantages for assessing aid effectiveness (Kilby, 2000). Since the project ratings data allows us to compare outcomes across different projects using the same evaluation criteria, it is by far

[12] This includes all education projects (projects that have more than 10 percent of education component) approved by education and non-education board between FY1996 and FY 2011. I selected the closed projects and excluded projects that are in "active," "pipeline," and "dropped" status since these projects do not have IEG ratings. In other words, the sample is limited to those projects that have been evaluated and received IEG ratings. Although Winters and Streitfeld (2013) randomly selected projects after dropping "pipeline" and "dropped" projects, I decided not to randomly select my sample as it will reduce my sample size. I discuss potential sample selection bias in the limitation section.

the largest and most consistent data that is available, and a standard base for the study of aid effectiveness (Kilby, 2000; Chauvet, Collier, & Duponchel, 2010). Also, the IEG data on project performance is the only data currently available with direct input and output relationships at the project level that can be compared across countries. Because of these advantages of the World Bank data, researchers in the past have utilized the data to study the importance of donors' efforts and recipients' macro-economic and institutional characteristics for the success of World Bank projects (Isham & Kaufmann, 1999; Dollar & Svensson, 2000; Kilby, 2000; Dollar & Levin, 2005; Chauvet, Collier, & Fuster, 2006).

For research question 1, I use data obtained from two publicly available sources. First, the IEG project performance ratings dataset is downloaded from the IEG website[13]. Once a project is completed, the World Bank prepares a completion report (ICR) and assigns ratings for all projects approved by the Board, and then IEG validates all of them and gives ratings (World Bank Independent Evaluation Group, 2012). The IEG dataset contains both the ICR and IEG ratings resulting from IEG's project assessments from the 1970s to 2011. I use the final IEG ratings as my dependent variable. The unit of observation of this dataset is a single project assessment, which is an independent validation of a completion report (ICR review), or an in-depth field based project evaluation (PPAR) (see Background section for details). In both cases, the same criteria are used to rate performances (World Bank Independent Evaluation Group, 2012). For any project that has earlier ratings, I use the most current ratings that supersede the earlier

[13] The World Bank continuously updates the IEG ratings data as more projects are completed and receive IEG ratings. I downloaded the IEG ratings data from the website that has up to date information as of January 2014.

ratings, as suggested by the IEG (World Bank Independent Evaluation Group, 2012). Education sector projects comprise about 20 percent of the whole IEG ratings dataset.

The IEG project performance dataset consist of basic information on each project in addition to its evaluation ratings. Each project has its project identification number, project name, borrowing country name, region as corresponding World Bank administrative geographical region, approval (the year when the Board approves lending operations) and exit year (the fiscal year the project actually closes), total project cost (this includes non-Bank funding), lending instruments (divided into two basic types: long-term focus specific investment, and short term focused development policy loans), and sector board.

The second dataset is obtained from the World Bank Projects and Operations website, where I downloaded the data on education projects which have detailed project information, such as subsectors, share of largest subsector and themes for each project. There are 1,694 projects that are categorized under "education" on the Bank website. The projects under the "education" category at the website includes projects from education sectors and projects from non- education sectors with education components. When downloading the information on the projects, I selected the closed projects, as on-going projects and cancelled projects will not have IEG evaluation ratings. I also excluded the projects that have less than 10 percent of education component, as those projects are more likely to be a non-education project with a very small education or training component. Using the project identification number and evaluation type[14], I merged the

[14] Since some projects had two types of IEG evaluations (PAR, PCR, EV, EM), the project ID was not a unique identifier. Therefore, I merged using the two variables project ID and IEG evaluation type, and then dropped the earlier evaluation type to have only one uniquely identified project. I dropped 240 education projects that had duplicated project ID number.

IEG project performance ratings dataset with the dataset downloaded from World Bank projects and operations website. Projects with outcomes of "Not rated" or "Not available" are dropped from the data.

The validity of these World Bank project ratings data may be questioned, and thus I address this concern empirically later in this study. However, regardless of whether the ratings are empirically proven to be valid later in this study, for this section, I follow the argument of Dollar and Svensson (2000, pp. 897–899) and Denizer et al. (2011, pp. 7–10) as they explain why the IEG outcomes are acceptable measures of program success. First, they argue that the IEG is independent of the Bank's senior management. This means that the IEG rating that is used in this study is an independent evaluation of the Bank management's self-assessment (ICR ratings), thus, there is less room for subjective evaluation of each project. Second, the outcome variables are highly correlated with improvements in observed economic performance, meaning that a project's "success" is not an irrelevant indicator of overall economic development of the country. Third, the IEG ratings are not significantly different from the more in-depth and detailed Project Performance Audit Reports. Since the PPAR is rated by evaluators that actually visit the project sites to assess the project results, it is more objective way to evaluate projects. In addition, some interviews of experienced IEG staff members supported the view that there is no certain evidence to believe the IEG ratings are biased (Smets et al., 2012). Statistically, a random measurement error in the dependent variable will not bias the coefficient but it can lead to larger standard errors (Wooldridge, 2010). The validity of the rating will not bias the explanatory variables that are used in the analysis, however, could lead to insignificant results of many variables.

Country level data sets

In addition to the World Bank data, country level data is merged with the World Bank dataset. Country level variables used in the analysis are obtained from two sources: 1) World Development Indicators 2) World Governance Indicators. The World Development Indicators are collected by the World Bank and compiled from officially recognized international sources (World Bank, 2015b). It is a time series data, available from 1960 to 2013. From this dataset, I choose to use measures of country's GDP growth, gross primary school enrollment, and spending in education as percent of GDP in the analysis, since these variables have small portion of missing data and strong initial correlation with the outcome variable.

I also merged the dataset with the Worldwide Governance Indicators that includes aggregate measures of six dimension of governance starting in 1996: voice and accountability, political stability and absence of violence/terrorism, government effectiveness, regulatory quality, rule of law, and control of corruption. The indicators are based on 31 different data sources, capturing governance perceptions as reported by survey respondents, non-governmental organizations, commercial business information providers, and public sector organizations worldwide (Kaufmann, Kraay, & Mastruzzi, 2010; World Bank IEG). Among these six indicators, I choose to use the government effectiveness measure from the six measures because of its strong initial association with the outcome variable. Each of the country level data is available for each year; therefore, I calculated an average value of the indicator for each project's duration using the approval and exit year information from the IEG data. Thus, the country level indicators differ from project to project, even within the same country.

The final dataset that was ready for analysis included 1,113 observations/projects from 127 countries. I use two kinds of datasets: one with project approval date from 1986 to 2011, the other one with project approval date from 1996 to 2011. This is because I include the government effectiveness variable in the model and the variable is measured after 1996. Thus, for the models that include the government effectiveness variable, I use the projects from 1996 to 2011. I discuss the findings mostly on the models that use this variable; therefore, the projects approved from 1996 to 2011 are the main data for this study. I describe more about the data and each of the variable in the section below.

Table 2 Descriptive statistics for all variables

		Mean	SD	Min	Max	Mean	SD	Min	Max
		Projects 1986–2011				Projects 1996–2011			
Project outcomes	Project Outcome (6 categories)	3.06	1.10	0	5	3.04	1.01	0	5
	Binary project ratings	0.74	0.44	0	1	0.73	0.44	0	1
Basic project characteristics	Project length	4.56	3.02	0	15	3.76	2.78	0	15
	Project cost (in USD millions, logged)	5.25	1.19	0	6	5.20	1.27	0	6
	Development policy loan (=1, 0=Specific investment lending)	0.25	0.43	0	1	0.32	0.47	0	1
	IDA (1=IDA, 0= IBRD)	0.62	0.49	0	1	0.61	0.49	0	1
	Education sector	0.51	0.50	0	1	0.45	0.50	0	1
	Repeater project	0.28	0.45	0	1	0.30	0.46	0	1
	Share of largest subsector (in %)	52.29	22.57	2	100	50.34	21.11	2	100
Loan Type	Emergency Recovery Loan	0.07	0.26	0	1	0.08	0.27	0	1
	Learning and Innovation Loan	0.02	0.15	0	1	0.03	0.17	0	1
	Specific Investment Loan	0.51	0.50	0	1	0.44	0.50	0	1
	Technical Assistance Loan	0.01	0.12	0	1	0.01	0.08	0	1
	Structural Adjustment Loan	0.05	0.22	0	1	0.06	0.24	0	1
Subsector	Primary	0.36	0.48	0	1	0.36	0.48	0	1
	Secondary	0.16	0.36	0	1	0.14	0.34	0	1
	Tertiary	0.16	0.36	0	1	0.11	0.31	0	1
	Vocational	0.06	0.24	0	1	0.04	0.20	0	1

	General education	0.18	0.39	0	1	0.20	0.40	0	1
Country characteristics	Primary school gross enrollment (USD millions, logged)	4.60	0.24	3	5	4.61	0.21	4	5
	GDP growth (annual %)	5.01	5.15	-8	46	5.15	5.64	-8	46
	Education spending as % of GDP	3.93	1.48	1	14	4.06	1.43	1	14
	Government effectiveness	-0.43	0.54	-2	1	-0.48	0.52	-2	1
Regions	Africa	0.27	0.44	0	1	0.26	0.44	0	1
	East Asia Pacific	0.09	0.29	0	1	0.07	0.26	0	1
	South Asia	0.13	0.33	0	1	0.09	0.29	0	1
	Europe and Central Asia	0.15	0.36	0	1	0.18	0.38	0	1
	Middle East and North Africa	0.05	0.21	0	1	0.03	0.18	0	1
	Latin America and Caribbean	0.32	0.47	0	1	0.36	0.48	0	1
	Observations	1261				976			

Note: Missing data for education spending variable is imputed

Variables

Outcome of interest: IEG project ratings

The outcome variable, the IEG project performance rating, is defined as the extent to which the project's objectives have been achieved efficiently (World Bank IEG, 2012). As explained in the background section, the IEG rates project outcomes using three criteria: relevance, efficacy, and efficiency. Based on these criteria, the current IEG rating system has a six-scale rating: highly unsatisfactory, unsatisfactory, moderately unsatisfactory, moderately satisfactory, satisfactory, and highly satisfactory. Keeping in mind that these ratings have been controversial (as explained in the background section), I use the IEG evaluation rating as my dependent variable for the quantitative analysis.

Each rating category has its own definition. For example, a highly satisfactory project is defined as "there were *no* shortcomings in the operation's achievement of its objectives, its efficiency, or in its relevance." A satisfactory project has the same definition with *minor* shortcomings, a moderately satisfactory project has *moderate* shortcomings, a moderately unsatisfactory project has *significant* shortcomings, an unsatisfactory project has *major* shortcomings, and a highly unsatisfactory project has *severe* shortcomings regarding the three criteria of objectives, efficiency or relevance (World Bank IEG). Although the definition of the each rating scale is somewhat subjective, overall, an evaluator's judgment about outcome is answering the question: Did the project achieve satisfactory development results, considering the importance and relevance of its major stated objectives and the associated costs and benefits? (World Bank Independent Evaluation Group, 2009). Therefore, a satisfactory project is a project that has fully delivered its objectives/goals, likely to sustain after the project is completed,

and show clear benefits in an efficient manner (World Bank Independent Evaluation Group, 2009).

It is important to note that the rating scale changed from binary (satisfactory/unsatisfactory) to the current six-scale rating in 1995. (Denizer, Kaufmann, & Kraay, 2013) uses two kinds 1) binary outcome and 2) sub-data set of ordinal outcomes. However, studies examining ratings in both categorical and binary forms generally do not find compelling reasons to use the six-scale ratings (Kilby, 2012). In fact, the bulk of the literature that used the IEG data has used the binary outcomes. Thus, my main results and discussion are based on the results with binary outcomes. Figure 8 shows that 74 percent of the projects in the data that were implemented between 1986 and 2011 received "satisfactory" ratings, when the six categorical outcomes are categorized into either "satisfactory" or "unsatisfactory."

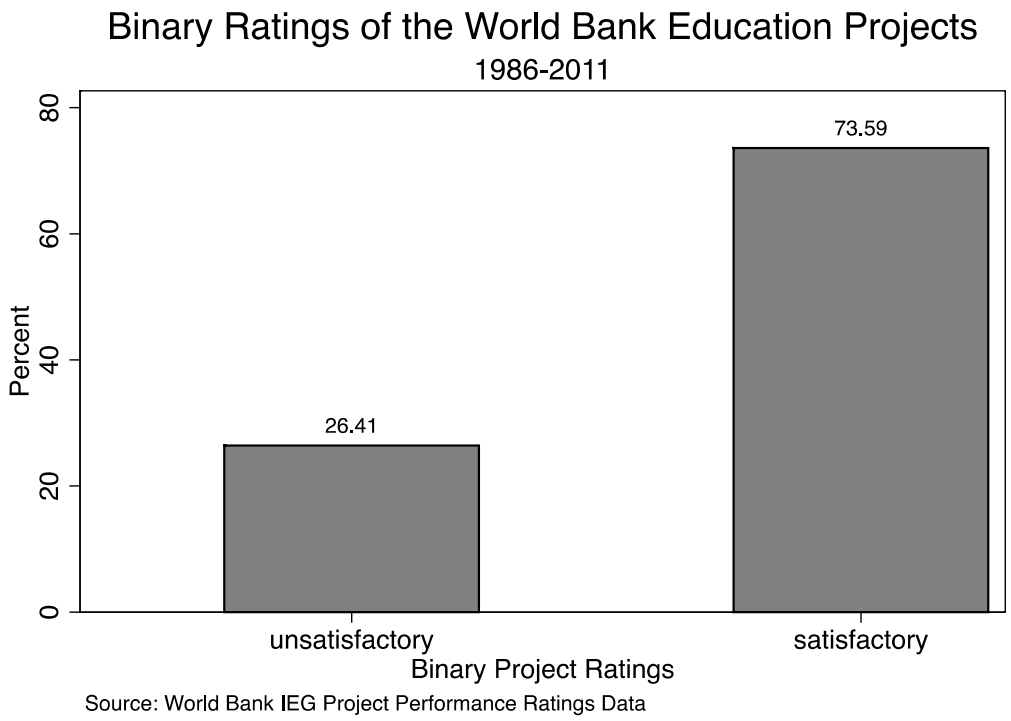


Figure 8 Binary ratings of the World Bank education projects (1986–2011)

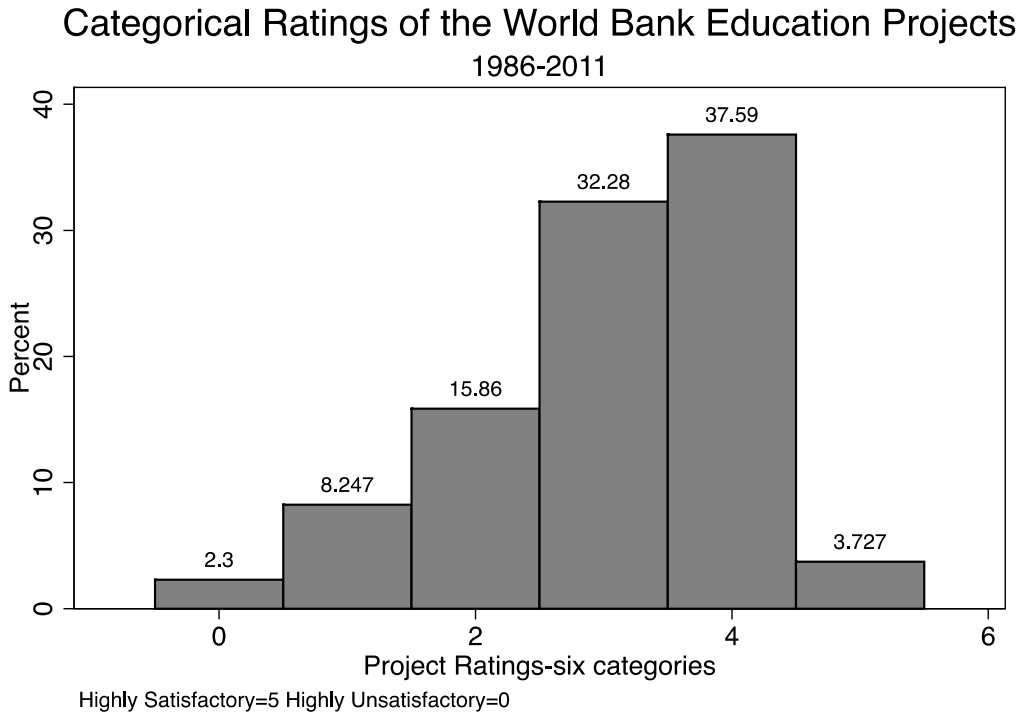


Figure 9 Categorical ratings of the World Bank education projects (1986–2011)

The differences in ratings between different time periods do not vary a lot, however, as Figure 10 shows, the ratings of the projects implemented in the 2010s are slightly lower than the projects implemented in the 1990s and 2000s. In the 1990s and 2000s, the largest proportion of projects received "satisfactory (coded as 4 in the six-categorical variable of outcome)," however, in the 2010s, the largest proportion of projects (37%) received "moderately satisfactory."

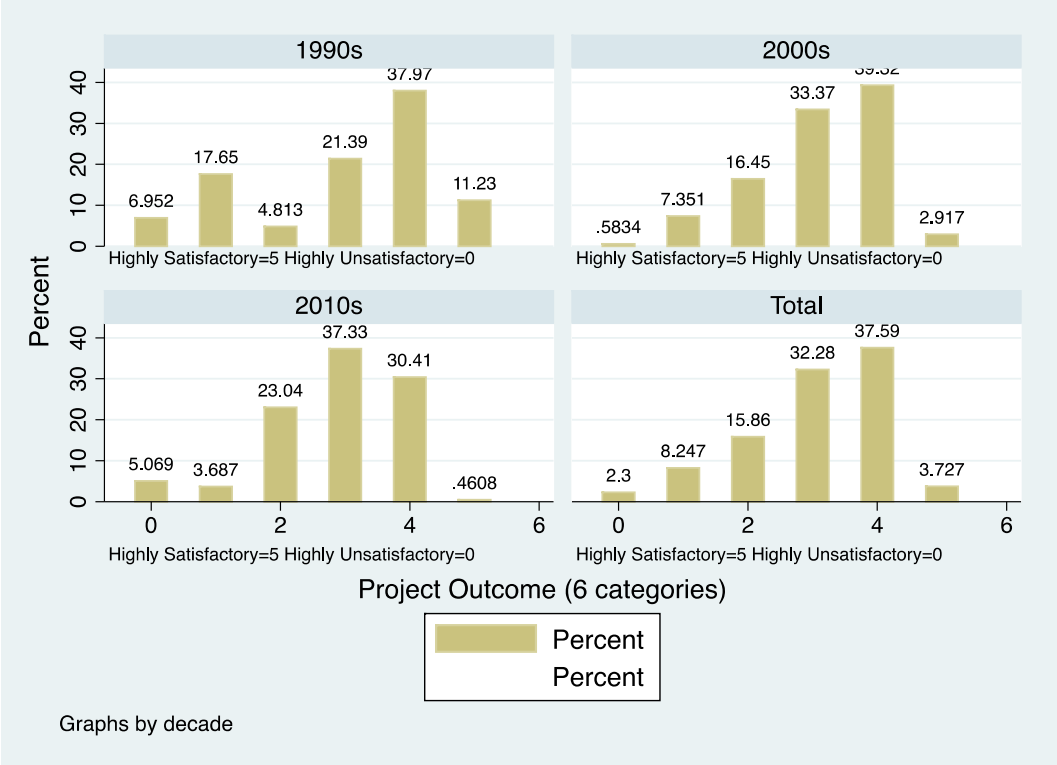


Figure 10 Project ratings by decades

The IEG project outcome data is derived from different types of evaluation reports. The IEG keeps a database for all completed World Bank projects, and the ratings are a mixture of ratings from the IEG Project Performance Audit Reports (PPARs), Project Completion Reports (PCRs), Evaluation Summaries (ES) and Evaluation

Memorandum ratings. The IEG conducts its separate evaluations for only about 25% of the total projects, and the outcomes of these evaluations is are the PPARs. For the rest of the projects, the IEG conducts desk reviews of the ICRs or PCRs completed by the task team leaders. PCRs are the ratings from ICRs that IEG kept in its database until the end of 1996 (Kilby, 2012). Starting in 1995, the IEG began to validate all ICRs and replaced old PCR ratings with new ratings (called PAR ratings) for the projects completed prior to FY 1983 (Kilby, 2012). The IEG also included an additional validation before the ratings, which is termed Evaluation Summary or Evaluation Memorandum ratings (Kilby, 2012; Denizer, Kaufmann, & Kraay, 2013). Since the performance ratings come from different evaluation sources, one might think it is necessary to control for evaluation type in regression model. However, since the initial correlation and the initial and final regression results between evaluation type and IEG ratings outcome did not show strong relationship between those two variables, I decide not to include the evaluation type in my model.

Another point to note is that some projects have revised their objectives and received higher ratings with the revised objectives compared to the original objectives. The IEG project outcomes data addressed this issue. The OPCS guideline (2006) describes that the overall rating is derived by weighting the rating of original objectives and revised objectives by the proportion of actual disbursement in each period, and rounded to the nearest whole number (1 to 6). Thus, I do not include an indicator for whether the project revised its objectives in the analysis.

Country-level variables

To add country level variables, I began by selecting a number of country level variables that have been identified in the literature as important determinants for IEG ratings, such as Country Policy and Institutional Assessment (CPIA), economic growth, measures of income inequality (e.g. Gini coefficient), measures of degree of political freedom (e.g. Freedom House index). However, not all of the measures and indicators used in other studies are used in my study, due to availability of the data, large portion of missing data, and their initial weak correlation with project outcomes. Moreover, some preliminary interviews I have conducted with the World Bank education staff indicated several important country level variables that I should include in my study. Based on the literature, the data availability, and the preliminary interviews with the education staff, the country level variables used in the final analysis are: GDP growth (annual percent), primary school gross enrollment, education spending as percent of GDP, and government effectiveness.

Since this study is about education projects, I add country characteristics that reflects educational context in each country: primary school gross enrollment and education spending. The primary school gross enrollment was included in the model in order to control for the overall size of primary education system. Primary school gross enrollment is a ratio of total number of students who are enrolled in primary school, regardless of their age, compared to the number of official primary school age (World Bank, 2015b). The gross enrollment ratio can exceed 100 percent as it includes those students who are over-aged (as well as under-aged) due to late enrollment or grade repetition, which is substantial in some countries particularly in Africa and Latin America

(World Bank, 2015b; Michaelowa & Weber, 2006). For this reason, its increase can indicate inefficiency rather than true increase in enrollment (Michaelowa & Weber, 2006). Thus, researchers tend to prefer net enrollment, which only considers students of the appropriate age. However, net enrollment data in developing countries are not widely available. Due to the limitation of data availability, some researchers used net enrollment data with limited scope of data, imputed the data using the information from gross enrollment and other indicators available, or have used gross enrollment ratio (Dreher, Nunnenkamp, & Thiele, 2008; Michaelowa & Weber, 2006; Michaelowa & Weber, 2008). Therefore, due to the wider availability of data and to prevent from loss of observations, I choose to use the gross enrollment ratio rather than net enrollment ratio, despite its impreciseness[15]. I use the logged primary school gross enrollment ratio. Although I hypothesize the increase in gross primary school enrollment is associated with high project ratings, it is possible that there could be negative relationship, if gross enrollment reflects inefficiency in the education system.

I also include education spending as percent of the GDP in the model. Spending on education variable measures public expenditure on education including government spending on educational institutions, education administration, and subsidies for private entities (World Bank, 2015b)[16]. Education spending as percent of GDP can reflect the extent to which the government values and prioritizes education. A country that appreciates and values education will invest a lot more money in education sector than a

[15] I tried including net enrollment ratio instead of gross enrollment ratio in the final model; however, the result was not much different from that of gross enrollment ratio. Thus, I use gross primary school enrollment ratio to prevent from losing more observations.

[16] The World Development Indicator also provide education spending as percent of government expenditure, however, this variable could not be used because of its large portion of missing data (31.3%). Education spending as percent of GDP also has missing data but slightly smaller than that of education spending as percent of government expenditure.

country that do not. A government that tends to spend more in education is likely to value education and thus prioritize implementing any education development projects that will lead to higher chances of project success. Thus, I hypothesize that an increase in education spending is associated with better project ratings.

Besides the two measures for education, I also add the annual GDP growth (measured in percent) in the model, to control for the level of economic development in each country. Previous studies have included GDP growth in predicting World Bank project outcomes, including Isham and Kaufmann (1999) and Denizer et al. (2013). GDP growth can also be a proxy for any macroeconomic shocks that may influence project outcomes (Denizer, Kaufmann, & Kraay, 2013).

It is extremely important to include GDP growth especially when education spending is included in the model. For example, since the 1960s, Sub Saharan Africa (SSA)'s public spending on education (in GDP terms) has compared favorably to that of successful Asian countries. The median SSA country spends a higher share of its budget (17.6 percent) and GDP (4.7 percent) on education than other developing regions (Fredriksen & Kagia, 2013), p. 277). However, their GDP per capita fell by 36 percent on average between 1970 and 1997 (World Bank, 2000a). Since SSA countries had much larger population growth while there was decline in GDP growth, the education expenditure growth is much less than that of East Asian countries that had high sustained economic growth per capita. Since the different levels of growth in public education spending are generated by very different levels of economic growth, I add the GDP growth in my model.

I include government effectiveness variable obtained from the Worldwide Governance Indicator dataset. Government effectiveness captures perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies (Kaufmann, Kraay, & Mastruzzi, 2010). This indicator is compiled by the World Bank Institute from nearly three dozen sources, including survey institutes, think tanks, non-governmental organizations, and international organizations (Moloney, 2009). It is a continuous variable with each country indicator a point estimate between 2.5 and -2.5. The highest value (2.5) is assigned to countries with the best governance while the lowest (-2.5) indicates the worst governance regimes (Kaufmann, Kraay, & Mastruzzi, 2006).

The Paris Declaration on Aid Effectiveness theoretically supports adding the government effectiveness variable. Hosted by the French government and organized by the OECD, a meeting that took place in Paris in 2005, where over 100 developed countries (the donors), developing country governments (the recipients), multilateral donor agencies, regional development banks, and international agencies endorsed an agreement to change the way they do the aid business more effectively. The Paris Declaration contains five principles, one of which is "ownership." The concept of "ownership" that is defined in Paris Declaration requires the developing countries to set their own development strategies, improve their institutions and tackle corruption (OECD). The concept of ownership in the Declaration emphasizes not only the leadership in developing national development strategies, but also taking the lead role in coordinating aid at all levels while maintaining dialogue with donors (OECD, 2005).

Since the borrower government implements the World Bank projects, the government is playing the leading role in coordinating the project, and thus, the project outcomes are likely to be affected by government effectiveness, or the quality of public services. I hypothesize that the increase of government effectiveness is associated with higher project ratings.

All of the country indicators in the dataset are measured each year. For example, the GDP growth measure is available from the World Development Indicator from 1986 to 2011 for each year. I take an average of each country measure for each of the project period. Thus, although I call it "country-level variable" it is actually measured and analyzed at the project level, since each project started and ended at a different year and thus the average value of country variables are different for each project.

Regional dummies are included in the model. Each project belongs to either one of the six World Bank administrative regions: Sub-Saharan Africa (AFR), East Asia and the Pacific (EAP), Europe and Central Asia (ECA), Latin America and Caribbean (LCR), Middle East and North Africa (MNA), and South Asia (SAR). Of the total education projects, 32 percent were implemented in the Latin America and the Caribbean region, while 27 percent were implemented in Africa. In East Asia and the Pacific region, 9 percent were implemented. Preliminary interviews indicated that projects implemented in Sub-Saharan Africa are in an environment where it is difficult to successfully implement projects. Figure 11 also shows that the Sub-Saharan Africa project ratings have the lowest portion of projects that received satisfactory ratings among all other regions. Only 62 percent of projects implemented in Africa were evaluated as "satisfactory." On the other

hand, the projects implemented in South Asia show the highest evaluation rating, as 91 percent of the all education projects were evaluated as “satisfactory.”

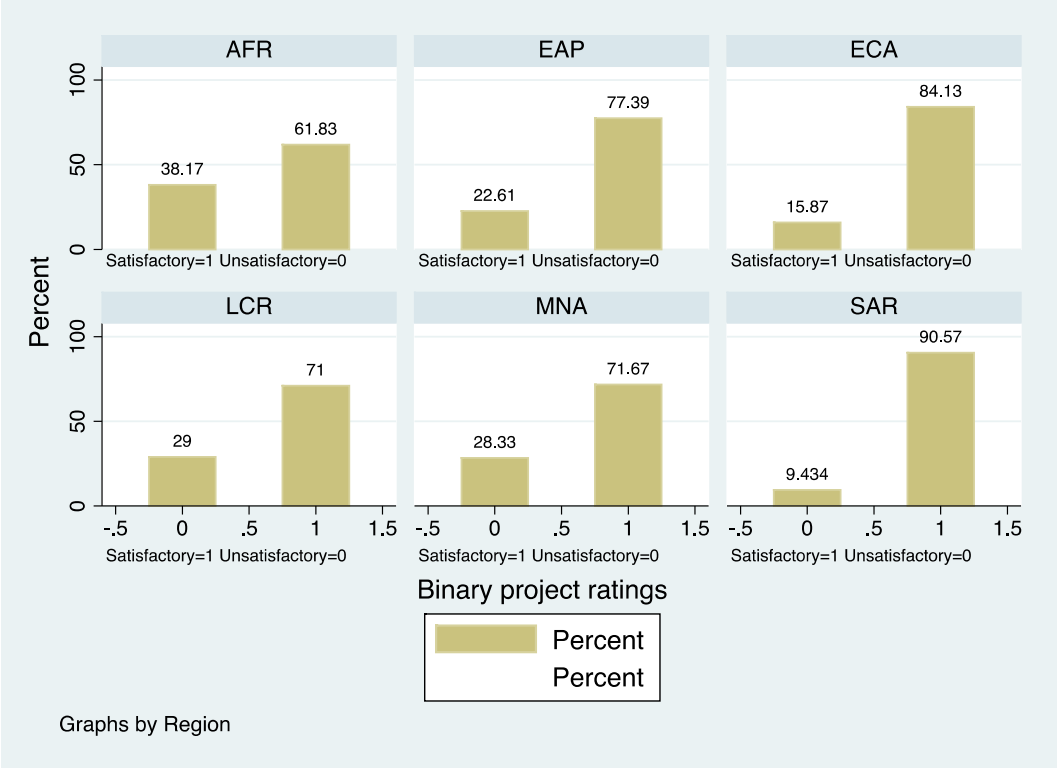


Figure 11 Project ratings by region

Project level variables

There are several groups of project-level variables I included in the model, as Denizer et al. (2013) found that the project level variables explain more variation in project outcomes than country level variables. I categorize project level variables into two types: 1) administrative factors 2) content related factors.

1) Administrative factors

I include basic information on each project in my analysis. First, the project length/duration variable is the length of time needed to implement the project, generated using the project approval and the exit fiscal year information from the World Bank dataset. The summary statistics in Table 2 shows that the average duration for education projects was 5 years (4.56 years), with a maximum of 15 years. The project cost is measured in units of USD millions, and is an estimated total cost of the project that includes non-Bank funding from other multilateral and bilateral agencies. I hypothesize the longer the project period, the increase of project cost is associated with higher project ratings.

As mentioned in the previous sections, the funding sources of the projects from the World Bank are either from IBRD or IDA. Funding from the World Bank is either from a type of loan provided by IBRD, or credit by IDA, to the borrower (World Bank Independent Evaluation Group, 2012). As the IDA provides loans and grants for the poorest countries, it is possible that the political and socio-economic environmental factors associated with project implementation in IDA countries are different from that of IBRD countries. Therefore, it is necessary to distinguish between IDA and IBRD countries in the model. Thus, I create a dummy variable for IDA. IDA here is not considered a country level variable, as there are some IDA borrower countries who "graduated" from IDA and became middle-income countries (IBRD borrowers). For example, a country who used to be a borrower for IDA loans in the 1990s could have turned into an IBRD borrower in 2000s. Projects implemented when the country is defined as IDA category were 62 percent of all education projects. I hypothesize that

projects implemented in IDA countries receive lower ratings than the projects in IBRD countries.

Another important variable to include in the model is the lending instrument type. As explained in the previous section, projects and operations are delivered in two types: investment and development policy loans. In my sample, 25 percent of the total education projects are development policy loans. Despite the small portion of development policy loans, given the different nature of these two types of lending instrument, I include a dummy variable for development policy loans in the model.

Although the lending instrument types are largely divided into two: investment loans and development policy loans, within the investment loans, there are several types. The database has seven different types of investment loans: adaptable program loan, financial intermediary loan, emergency recovery loan, sector investment and maintenance loan, specific investment loan, learning and innovation loan, and technical assistance loan. However, the majority of investment loans are specific investment loan (51 percent). Development policy loan replaces the previous different types of lending (e.g. structural adjustment loans, sector adjustment loans, and poverty reduction strategy credit). Figure 12 below shows the project outcomes depending on types of lending. Figure 12 shows not much variation of outcomes depending on the loan types. However, as the project activities and characteristics are different based on loan types, I include the dummy for each loan type in my model. After looking at the initial correlation with the project outcomes with each of these loans, I selected the following loans to include in the model: emergency recovery loan, learning and innovation loan, specific investment loan, technical assistance loan and structural adjustment loan.

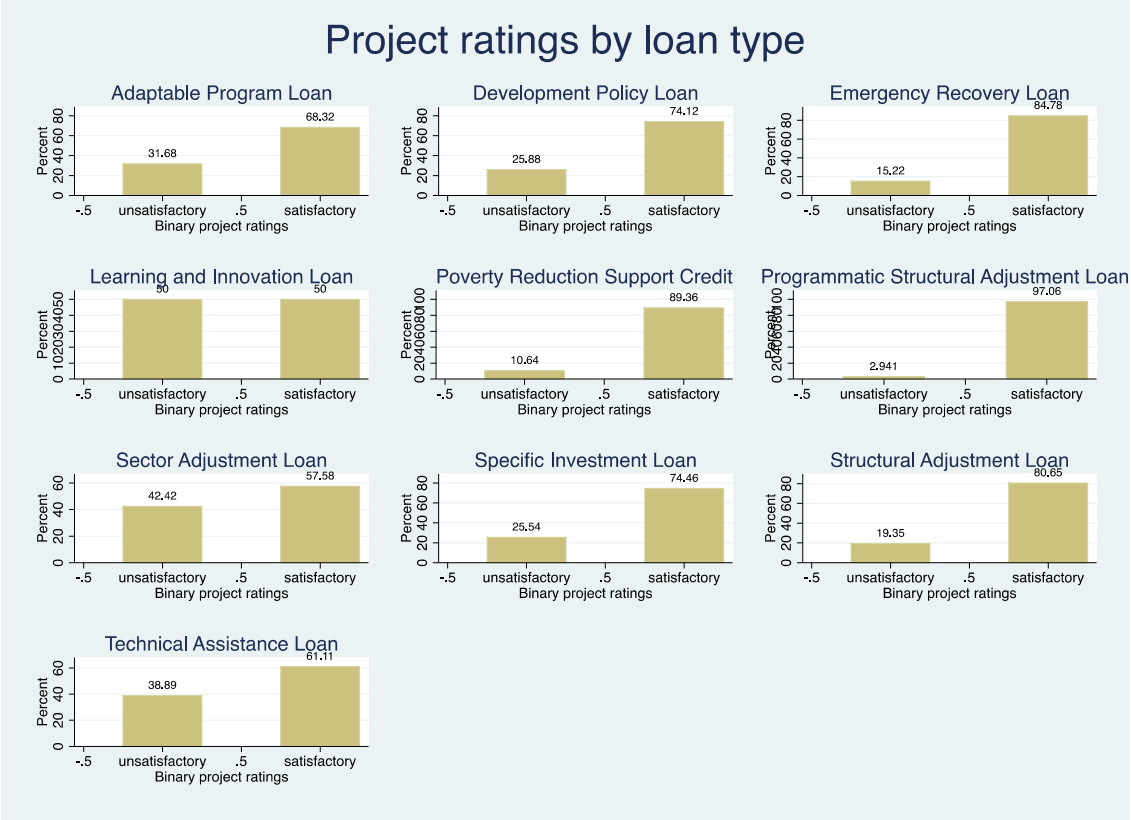


Figure 12 Project ratings by loan type

Emergency recovery loan is for social and economic recovery immediately after an extraordinary event such as war and natural disaster (World Bank Independent Evaluation Group, 2012). Figure 12 shows that emergency recovery loans have a high proportion of satisfactory ratings. Since emergency recovery loan can have a larger impact than other loan projects in countries without any disaster, I expect a positive relationship between emergency recovery loan and project outcomes. Learning and innovation loan are usually loans of $5 million or less aimed to finance small, experimental, risky and time sensitive projects in order to pilot initiatives or experiment a locally based models prior to a larger-scale intervention (World Bank Independent Evaluation Group, 2012). Given that small-scale projects have less chances of meeting

implementation challenges, it can be expected that learning and innovation loan is associated with better project outcomes. However, the Figure 12 shows that about half of them have received unsatisfactory ratings.

Specific investment loan is the most common loan type of education project, and in my data about half of total projects are specific investment loans. It supports the creation and maintenance of economic, social and institutional infrastructure. Technical assistance loan focuses on organizational arrangements, staffing methods, and technical, physical or financial resources in key agencies. I have also added a former type of development policy loan, a structural adjustment loan in the model. Structural adjustment loan promotes policy reforms rather than financing investment in developing countries, and often carried out with conditionalities that the World Bank imposes to the government. These two types of loans also have a high proportion of unsatisfactory ratings (Figure 12). One of the most important reasons for this may be the difficult political economy of implementing reforms in contracting economies.

2) Content related factors

Since the projects include projects in education sector and the projects with education components in other sector, I create a dummy variable to determine if the project belongs to the education sector (education = 1, other sector =0). The proportion of projects that belonged to the education sector board was 51 percent. Almost half the projects with education components were implemented in other sectors. Table 3 below shows which sector the other half of the projects with education components belongs to. Many projects were implemented in the social protection sector and public sector

governance, beside education sector. I expect a project that belongs to education sector to have a positive relationship with project outcomes.

Table 3 Projects with education components in education and non-education sectors

Sector board	Number of projects	Percentage of total
Agriculture and rural development	35	2.8
Economic policy	93	7.4
Education	641	50.8
Energy and mining	2	0.2
Environment	8	0.6
Financial sector	2	0.2
Financial and private sector development	16	1.3
Health, nutrition and population	26	2.1
Poverty reduction	41	3.3
Public sector governance	142	11.3
Social development	12	1
Social protection	208	16.5
Transport	4	0.3
Urban development	32	2.5

I also create a set of dummy variables for subsectors within the education sector. Education projects can fall into many different categories of subsectors, and I create a set of dummy variables for primary, secondary, tertiary, vocational and general education subsectors[17].

[17] These subsector dummies are not mutually exclusive categories, as a project can have many components from different sectors. For example, a project can have 20 percent of the component in primary education sector and the rest in secondary or vocational education sector. The data has information on the proportion of each of the subsector components. Based on this information, I created dummy variables for subsectors when the proportion of subsector component is larger than 20 percent. I also tried several different threshold (e.g. 30%, 50%) for creating subsector dummies, however, the result were not much different from the 20% threshold that I used in the final analysis.
General education subsector indicates programs or activities implemented at education system level rather than one educational level. Activities at general education subsector can be training of Ministry officials, supporting decentralization or monitoring and evaluation, and improvements of Education Management Information Systems (EMIS). It focuses on World Bank operations on increasing government's capacity to deliver quality education services (World Bank, 2010).

Projects that are focused on primary education are the majority of all projects in the data (36%). Other subsectors have a very small number of projects. Figure 13 shows that there isn't much variation in project outcomes depending on project subsector. Figure 13 shows that the largest proportion of projects that received "satisfactory" ratings is secondary education projects (79.7%), and the smallest in tertiary education sector (72%), however, the difference between the proportions is very small, only 7.7 percentage points. However, theoretically, distinguishing project subsectors is important, as project characteristics are different in each subsector. Therefore, I include this variable in the model.

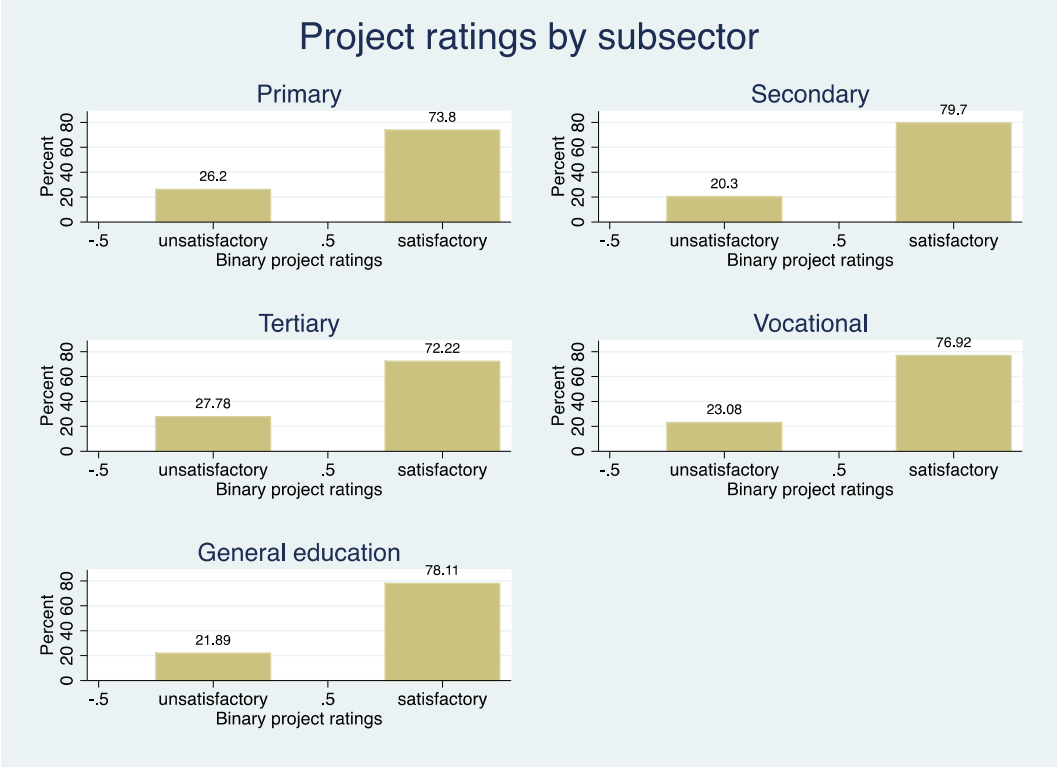


Figure 13 Project ratings by subsector

I also create a dummy variable for a project that is a follow up of previous projects, and is in the second or third phase of initial project (e.g. Primary education development program II, III). Any project name with roman numeral II or III is considered a "repeater" project. To identify such projects, I searched the names of all projects for numbers and roman numerals, and created a dummy variable equal to one if a project is the second or higher of an initial project, and zero otherwise. The descriptive statistics table (Table 2) also shows that 28 percent of the projects were a "repeater" projects, meaning that they are the second or third phase of a project that was implemented prior to the project. I expect that a project in the continuous stage (second or third phase projects) receives better ratings.

Lastly, the percent share of largest subsector is added in the model. Denizer, Kaufmann, & Kraay (2013), who also looked at this variable in their study, as one of their dimensions of "project complexity." If the largest share of the project is assigned to a single subsector, meaning that higher percent is allocated to primary education for example, then the project is considered as less complex as it will have objectives focused on primary sector instead of objectives that spans multiple sectors. The average share of largest subsector is 52.3 percent of all education projects, with minimum two percent and maximum 100 percent. Based on preliminary interviews, I expect that a project with a larger portion of a subsector is associated with higher ratings, as it indicates that the project is less complex and has fewer objectives.

Missing data

The sample has some missing data in many of the variables that I include in the model. Table 4 shows the number and percent of missing values for the variables that have missing values.

Table 4 Missing data

Variable	Missing	Percentage missing
Project duration	1/742	0.13
Dummy for IDA	1/742	0.13
Primary school gross enrollment	47/742	6.33
GDP growth	4/742	0.54
Education spending as percent of GDP	127/742	17.12

Table 4 shows that the education spending variable has the largest portion of missing values among all the variables (17.12%). Therefore, I impute the missing values for the education spending variable only, using the multiple imputation method[18]. I assume that the missing data are Missing At Random (MAR), i.e. missing values do not carry any extra information about why they are missing than what is already available in the observed data. Under MAR, the probability of a value being missing depends on the observed values. Multiple imputation replaces the missing values with multiple sets of simulated values to complete the data, applies standard analyses to each completed

[18] Using the Stata 13 mi impute command, I created 10 imputation datasets and 1000 burn in iterations. Truncated normal regression was used as a method to impute the values for missing data for education spending as percent of GDP variable, as it should not have a negative value.

dataset, and adjusts the obtained parameter estimates for missing-data uncertainty (Rubin, 1987, p. 76). The objective of missing data is not to predict missing values as close as possible to the true ones but to handle missing data in a way resulting in a valid statistical inference (Rubin, 1996). Thus, multiple imputation can be more efficient than a commonly used listwise deletion (complete-case analysis) and can correct for potential bias. Having imputed the education spending variable using multiple imputation, all my main analyses are based on the imputed dataset.

Method

Potential sources of bias and analytical strategy

Unobserved country level factors

In order to check if there is a significant relationship between the project characteristics and project outcomes, I ran linear probability and logit models for binary project outcomes (satisfactory/unsatisfactory), and ordered logit models for the six categorical project outcomes (highly satisfactory to highly unsatisfactory). The partial correlation between the set of variables that measures project characteristics and project outcomes could be confounded by other unobserved variables that might affect both project characteristics and project outcomes. Serious confounding factors are related to country characteristics. For example, a country that prioritizes education is more likely to allocate the national expenditure on education and is also more likely to have successful project implementation. Other external factors, such as corruption, level of economic development, and political stability can influence both the process of implementing projects and project outcomes.

To address the issue of country heterogeneity, I control for some factors that reflect the government effectiveness in public policy and economic development (GDP growth). Considering the clustered nature of the data (projects implemented within countries), I use a cluster adjusted robust standard errors for all model specifications[19]. However, there could still be unobserved country heterogeneity that affects both the explanatory variables and project outcomes. Therefore, in addition to regular regression models, I add country fixed effects in OLS and logit estimator, which only observe the variation within countries.

Time trends are another important source of confounding factors. The World Bank changed its education lending policy over time, and this should have had a strong influence in identifying and preparing education projects. Macroeconomic factors, such as financial crisis, inflation or change in exchange rate, could also obstruct the implementation of projects, affecting project outcomes. Natural phenomena are another factor that could affect the project outcome to a large extent. For example, as described in the previous section, the 1972 earthquake in Nicaragua devastated the project that was being implemented at that time (Jones, 1992). Natural disaster can affect both macroeconomic factors and project level variables (e.g. increased cost or higher project

[19] Hierarchical linear modeling (HLM) can be used in this study, however I decide not to apply HLM model at this time. The advantages of using an HLM model are that 1) it can correct for standard errors that might be biased due to clustering, 2) it allows me to assume country heterogeneity by allowing random intercept, and 3) it allows me to investigate relationship between variables at different hierarchical levels (e.g. interaction between project and country level variable) more efficiently. However, I decide not to use HLM in this study for the following reasons. First, I can correct for standard errors for the nested data by applying cluster robust standard errors for all model specifications. Secondly, I can employ country fixed effects model to account for differences between countries. An HLM model does not correct the estimated impacts for project-level variables for any bias caused by unobserved country-level variables. Since I regard the threat to the validity coming from bias due to unobserved country heterogeneity is larger, I prefer to use country fixed effects model rather than an HLM model. Thirdly, while in HLM the relationship between a project and country can be explored, based on the initial results of my study, there was no significant project level variable. Therefore, I decide not to employ HLM. Future studies can employ an HLM if the research question is centered on the relationship between a project and country-level variable.

cost) as well as project outcomes. To address the concerns of World Bank policy change, external macroeconomic shocks, natural disasters or anything that happened in a specific year that could affect both the explanatory variables and project outcomes, I add year dummies for each year of the project implementation period in all model specifications.

Unobserved project level factors

Unobserved project level factors could also affect some of the explanatory variables as well as project outcomes, confounding the effect of the explanatory variable on project outcomes. For example, a project that is more challenging than other projects, and has many objectives and donors for financing, would usually take longer to initiate. At the same time, such a project is more likely to fail because there is higher risk in donor cooperation. If one of the many donors for the project decide not to cooperate during the implementation of the project, then the project is less likely to achieve its goals thereby leading to project failure. In this example, unless the levels of challenge of projects are considered in the model, it is likely that the project duration (length of time) will suffer from omitted variable bias.

The potential endogeneity of project variables have been recognized by other researchers. Deininger, Squire and Basu (1998) and Dollar and Svensson (2000) proposed various country and project characteristics as instruments for supervision costs. Kilby (2000) used lagged supervision and change of project performance over time. Kilby (2012) used country-level measures of political influence of donors on recipients as instrument for preparation time in examining the effect of preparation time on project outcomes. However, there are concerns about the justification of validity of the exclusion

restriction (which requires that the instrument, political influence, are affecting the project outcome only through project preparation time) of instruments used in these studies (Denizer, Kaufmann, & Kraay, 2013).

Moreover, there is almost no study that examined the World Bank project performance only in the education sector. This makes it difficult to find an instrument while so little is known about the mechanism and characteristics of education projects. In addition, the World Bank project performance data carries only information on administrative characteristics of each project, making it more difficult to add control variables or find an instrumental variable. Therefore, while I control for as many project level variables as I can, there are limitations of the econometric method in this study. Thus, this study attempts to first lay the foundation for understanding the mechanism and characteristics of education projects by further conducting qualitative interviews with the education sector staff at the World Bank.

Linear probability model

Keeping in mind the potential sources of bias, I start with a linear probability model, in which the coefficients are estimated by ordinary least squares (OLS). A linear probability model assumes a linear relationship between independent and dependent variables. OLS seeks to minimize the sum of squared distances of the data points to the linear regression line. Assuming zero conditional mean of the errors, I set up my basic model as below.

$$\Pr(y=1|x) = \beta_0 + \beta_1 Country_{ic} + \beta_2 Project_{ic} + e_{ic} \qquad (1)$$

Where,

Pr(y=1|x) is the probability of receiving a satisfactory rating for project *i* in country *c*, *Country* is a set of country characteristics including primary school gross enrollment, GDP growth, education spending as percent of GDP, and government effectiveness. *Project* is a set of project characteristics including project duration, cost, lending instrument, dummy for education sector, a set of dummy variables for each subsector, type of loans, dummy for repeater project, and share of largest subsector. *e* is the project specific error term. Based on this basic model, the full model (final specification) includes a set of regional dummies and a set of year dummies to account for regional differences and year trends. The final specification also includes some interaction terms.

However, the linear probability model has some drawbacks. It produces predicted probabilities that are less than zero or larger than one, while the probabilities must be between zero and one. Also, a probability cannot be linearly related to the independent variables for all their possible values (Wooldridge, 2009, p. 249). Thus if the majority of data is far away from the mean (either near zero or near one), this model would not be an efficient estimator. Another drawback of this model is that it contains heteroskedasticity, indicating that it can cause the standard errors to be biased, thus invalidating the test statistics. However, the OLS estimator is still asymptotically normal even if it contains heteroskedasticity, if the sample size is large enough (Wooldridge, 2009). I still conduct a White test to detect if the heteroskedasticity is a severe issue, however, did not find any clear evidence that heteroskedasticity is present.

Logit model

The drawbacks of the linear probability model can be overcome by using more sophisticated binary response models (Wooldridge, 2009, p. 575). A binary logistic regression can predict the logit of a project receiving satisfactory ratings by a set of predictors. The logit is the natural log of the odds, or probability/(1-probability) (Peng & So, 2002). The basic logit model I use is expressed as follows:

$$\Pr(y=1|x) = P(y^* > 0|x) = G(\beta_0 + \beta_1 Country_{ic} + \beta_2 Project_{ic}) \qquad (2)$$

Where G is a function taking on values between zero and one: $0 < G(z) < 1$, for all real numbers z. This ensures that the estimated response probabilities are strictly between zero and one, thus overcoming the drawback of the linear probability model (Woodridge, 2009). Since it is a logit model, G is the logistic function:

$G(z) = \exp(z)/ [1+\exp(z)] = \Lambda (z)$

Which is between zero and one for all real numbers z. This is the cumulative distribution function for a standard logistic random variable (Woodridge, 2009).

The logit model is derived from the underlying latent variable model. Because we can only observe project outcome as either one or zero, the project outcome, or the project "success," is observed indirectly using a latent variable (y*) that is predicted by a set of factors and a disturbance process ε.

$$y^* = \beta_0 + \beta_1 Country_{ic} + \beta_2 Project_{ic} + e, \qquad (3)$$

$y = 1[y^* > 0]$

The idea of a latent y* is that an underlying propensity for a project to become "successful" generates the observed state (Long & Freese, 2006). The notation 1[•] defines that y is one if $y^* > 0$, and y is zero if $y^* \leq 0$. The error term is assumed to be independent of the independent variables and has the standard logistic distribution, leading to the binary logit model (Woodridge, 2009; Long & Freese, 2006).

Whereas linear probability model is estimated by OLS, this is not applicable with the nonlinear nature of the logit model. The logit model is estimated by maximum likelihood estimation (MLE). The MLE seeks to maximize the log likelihood, which is how likely it is (the odds) to observe the project outcomes predicted by an explanatory variable. The general theory of maximum likelihood estimation for random samples is that under general conditions, the MLE is consistent, asymptotically normal, and asymptotically efficient (Wooldridge, 2009).

There is no universally accepted goodness of fit measure, therefore, to measure the goodness of the fit of the logit model, I examine several measures: percentage correctly predicted, McFadden's Pseudo-R squared, and area under ROC (receiver operating characteristic) curve. As mentioned before, y=1 if the predicted probability is at least 0.5, and zero otherwise. The percentage correctly predicted measures how well y=1 predicts every y across all observations. However, this measure can be misleading because it can have high percentages of correctly predicted even when the least likely outcome is very poorly predicted (Wooldridge, 2009).

I also look at the pseudo R squared for goodness of fit in the logit model. The pseudo-R squared captures the relative changes in the log-likelihood given the model parameters. I choose to look at McFadden's (1974) R squared as it is more intuitive,

upper bound is 1, and is closely related to R squared definition for linear models. McFadden's R squared is based on the calculation of mean of the predicted probabilities of an event for each of the two categories of project outcome variable, and the difference between the two means (Allison, 2013).

Another common measure used to evaluate the fit of a logistic regression model is the area under the receiver operator characteristic (ROC) curve. The goal of logistic regression is to generate an equation that can classify observations into one of two outcomes (Peng & So, 2002). The ROC curve measures the sensitivity (true positive) and specificity (true negative) versus predicted probabilities and it graphically shows the degree to which predictions agree with the data. The ROC curve is a plot of sensitivity (y-axis) versus 1 minus specificity (x-axis). Sensitivity is the proportion of observations correctly classified as an event of receiving satisfactory ratings (also called the true positive fraction). One minus specificity is the proportion of observations misclassified as an event of receiving satisfactory ratings (also called false positive fraction) (Peng & So, 2002). Hence, a model with a larger area below the ROC curve, the one associated with the greatest sensitivity and the lowest one minus specificity, is considered a better model (Afifi & Clark, 1990); (Peng & So, 2002).

In addition to the model fit, I also present the coefficients of logit models in the odds ratio, which is basically the effect sizes (Nandy, 2012). The odds ratios (OR) are nonlinear transformations of the logit coefficient, and represent the odds of the occurrence of an outcome given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure (Szumilas, 2010). Thus, the odds ratios are used to compare the relative odds of the occurrence of the outcome of interest

given exposure to the variable of interest (Szumilas, 2010). Positive effects are greater than one and negative effects are between zero and one (Long and Freese, 2006).

I also present graphs of predicted probabilities as the logit model is nonlinear. Predicted probabilities are outcomes when the values of independent variables are given. For example, with the predicted probabilities, I can look at the relationship between share of the largest subsector and project outcomes when the government effectiveness is low (defining it as one standard deviation below the mean), middle (at the mean), and high (one standard deviation above the mean). Because the results of predicted probabilities are contingent on specified values for all of the independent variables, I can set the rest of the independent variables at the mean values[20] (Long and Freese, 2006).

In sum, the logit model can be a more efficient estimator than the linear probability model because its nonlinear nature of the model ensures that the probabilities are between zero and one, overcoming the drawback of the linear probability model. Because of the nonlinearity of the logit model, maximum likelihood estimation is used to estimate the logit model. Unlike the linear probability model, logistic regression provides no universal good-of-fit measures, therefore I report several measures for the model fit. The logit model also provides odds ratio and well as predicted probabilities for more intuitive interpretation of the coefficients.

[20] For calculating the predicted probabilities, I do not use the imputed dataset as the predict command in Stata did not work after multiple imputation command (mi impute). Therefore, the results of predicted probabilities are based on the data set that is not imputed.

Ordinal logit model

In addition to the linear probability and logit models, I apply an ordered logit model to examine the determinants of education project outcomes. As explained previously, the IEG rating scale was changed to six categories in 1995. The scale of the project outcomes ranged from highly satisfactory to highly unsatisfactory, making it an ordered dependent variable. Although the ordinal variables are ordered/ranked, the distances between the categories are unknown, therefore it is better to avoid the assumptions that the distances between the categories are equal (Long and Freese, 2006). For example, the interviews with the Bank staff (Interview #4) indicated that it is very difficult to have a highly satisfactory project, whereas satisfactory or moderately satisfactory projects are very common. This implies that the distances between the rating categories may not be equal for IEG ratings.

An ordered logit model makes full use of the ordered/ranked data. The model preserves the information that the categories of the dependent variable are ordered, thus it is a better estimator than the binary logit or multinomial logit models, if the dependent variables are ordered outcomes. The ordered logit model considers a set of dichotomies, one for each possible cut off of the response. This is good for the parsimony of the model, because it means that the effect of an explanatory variable on the ordinal response is described by one parameter. Since there are six categories of project outcomes, 5 (6–1) cumulative probabilities are considered. The ordered logit model assumes the same coefficients for each cut off j. There is thus only one set of regression coefficients, as opposed to five as in a multinomial logit model.

The ordinal regression model can be represented as a latent-variable model, like the logistic regression model. The latent dependent variable, which is continuous, ranges from negative infinite to positive infinite. The latent variable can be thought of as the propensity to receive highly satisfactory, satisfactory or other categories of project outcomes. The observed project outcomes categories depend on the latent variable and the cut points for each category. In this study, the observed project *i*'s outcome looks as below.

$y_i^* = \alpha + \beta x_i + \varepsilon_i$ (4)

Where *i* is the project, *x* is a set of independent variables (country and project characteristics, year and regional dummies) and ε is a random error. The latent dependent variable *y** is divided into six ordinal categories according to the cut points as below.

$y_i =$

1 => highly unsatisfactory if $\tau_0 = -\infty \leq y_i^* < \tau_1$

2 => unsatisfactory if $\tau_1 = \tau_1 \leq y_i^* < \tau_2$

3 => moderately unsatisfactory if $\tau_2 = \tau_2 \leq y_i^* < \tau_3$

4 => moderately satisfactory if $\tau_3 = \tau_3 \leq y_i^* < \tau_4$

5=> satisfactory if $\tau_4 = \tau_4 \leq y_i^* < \tau_5$

6=> highly satisfactory if $\tau_5 = \tau_5 \leq y_i^* < \tau_6 = \infty$

Thus when the latent *y** crosses a cut point, the observed category changes. In other words, the value of the observed dependent variable *y* depends on whether it has crossed the particular threshold (Williams, 2015).

In sum, I use the ordered logit model because it will not lose the information contained in the discrete ordered variable. As explained in the binary logit model, it will

ensure that the observed project outcomes depend on the probabilities between the cut points. Because the distances between one category and another category (e.g. satisfactory to highly satisfactory) may not always be constant, and ordered logit model for ordered project outcomes should be more efficient than the binary logit models with binary outcomes. However, as mentioned in the section on dependent variable, the previous studies did not find a strong reason to use the six-scale ratings, therefore many studies that use the IEG data have conducted their analyses with binary outcomes (Kilby, 2012). Therefore, although I present the results from ordered logistic regressions, my main analysis is based on the results from binary logit models.

All model specifications (except fixed effects estimations) include standard errors clustered at the country level. Since projects are implemented within countries, and it is likely that projects implemented in the same country will have similar characteristics. Standard errors clustered at the country level in model specifications without country fixed effects takes into account the clustered IEG data. Year dummies are also included in all specifications to account for any external political economic shock or natural disasters that occurred during a specific year that might have largely influenced the project outcomes.

Linear probability model and logit model with fixed effects

Although I have controlled for many project and country level variables in the previous OLS, logit and ordinal logit models, there are other omitted factors that could bias the estimates of the set of observed country and project characteristics on project outcomes. In fact, it is clear that unobserved country effects (e.g. cultural differences) can

be an important factor in determining project outcomes. For example, if a country traditionally values education then it is likely that the government prioritizes education projects and proceeds faster in implementation, which makes the project outcomes likely to be significantly better than other countries. However, to focus on project level factors and time-variant country characteristics, I use a logit model with country fixed effects.

Fixed effects estimates use only within-country differences, essentially discarding any information about the differences between countries. If predictor variables vary greatly across countries but have little variation, fixed effects methods help control for omitted variable bias by having countries serve as their own controls (Allison, 2009). By including fixed effects, I can control for the average differences across countries in any observable and unobservable predictors.

Fixed effects are usually used in balanced or unbalanced panel data; however, the World Bank evaluation rating data used in this study is a clustered data set, which is a cross-sectional data, but each project belongs to a country (cluster) and does not involve time. Although the time period of project ranges from 1986 to 2011, since each project lasts for several years, the outcome of the project is available for only a specific year (usually after 3–4 years when the evaluation is being done after the project is closed). Fixed effects methods can be applied to a cluster sample, and it is preferred when we think the unobserved cluster effect is correlated with one or more of the explanatory variables (Wooldridge, 2009). Since the IEG evaluation ratings data is unbalanced, meaning that there are different numbers of projects within countries, fixed effects for unbalanced panels are required.

Fixed effects estimator, also called within estimator, subtracts the mean values of the country for each of the project implemented within the country (mean-differencing). The within estimator performs OLS or conditional logit on the mean-differenced data (Cameron & Trivedi, 2010). By subtracting the means, I am restricting all the actions in the regression to within-country variation, having eliminated the omitted variable bias, which are the unobserved across-country differences. By controlling for the average differences across countries in any observable or unobservable predictors, the fixed effects coefficients soak up all the across-country differences, leaving me with only within-country variation. Thus, the fixed effects estimator greatly reduces the threat of omitted variable bias due to unobserved country heterogeneity.

I first start the initial analysis by using an OLS fixed effects estimator as follows:

$$Project\ outcome\ _{ic} = \beta_0 + \beta_1 Country\ _{ic} + \beta_2 Project\ _{ic} + u_c + e_{ic} \qquad (5)$$

Where *project outcome* represents the probability of receiving satisfactory rating for project i in country c, *country* is a vector of country characteristics that are averaged for each project, *project* is a vector of project characteristics, *u* is a vector of country fixed effects (unobserved country factors that do not change over time for the same country) and e is idiosyncratic error term. The coefficient (beta) for each of the explanatory variables in this model indicates the average effect of the predictor, which is also the common slope averaged across the countries. The country variables in the model vary across projects, not across countries, as they are average value of each country indicator over the project period years. Therefore, applying country fixed effects estimator does not

eliminate the country variables as the mean-difference of the country variable will not be zero. In addition to OLS, I also apply country fixed effects with logistic regression models (conditional logit model). As shown in the equation, in a logit model with fixed effects, individual intercepts are created instead of fixed constants.

$$\Pr(Y_{ic}=1) = \exp(\alpha_i + x_{ic}\beta) / 1+\exp(\alpha_i + x_{ic}\beta) \qquad (6)$$

Because fixed effects rely on within-country and within-year variation, the estimator automatically drops the countries with only one project implemented within a year, therefore reducing my sample size[21]. I would also need a reasonable amount of variation of explanatory variables within each country. One limitation of fixed effects models is that I cannot assess the effect of variables that have little within-country variation. For example, if I wanted to know the effect of any country level predictors such as GDP growth on probability to receive satisfactory ratings, the fixed effects model is not appropriate because country level predictors do not vary a lot within countries. Therefore, in the fixed effects model, I pay more attention to the project level characteristics rather than country variables. Allison (2009) notes that in applications where the within-country variation is small relative to the between-country variation, the standard errors of the fixed effects coefficients may be too large to tolerate. However, a Hausman test strongly rejects the null hypothesis (Chi squared (27) = 414.72, $p<0.0000$) that the difference between fixed and random effects coefficients are not systematic,

[21] I acknowledge that using fixed effects, since the estimator drops countries with small number of projects, weakens the external validity while strengthening internal validity.

therefore, I conclude random effects estimator is inconsistent and continue with fixed effects estimation.

Limitations

This study is not without its methodological limitations. Although I controlled for unobserved country heterogeneity, there could still be other country characteristics that vary within countries that might have significant influence on the project ratings. For example, the estimator cannot capture the confounding factors in a country where the government is frequently changed within a year. In the previous section, I described that in Sudan, the government frequently changed, which impeded the implementation of projects (Jones, 1992). In order to overcome this problem, Kilby (2012) uses government fixed effect instead of country fixed effect. However, as government variable is not obtained; I am not able to use government fixed effect.

Other unobserved factors at the project level can also lead to potentially biased results due to omitted variable bias. One example can be the quality of the project task team leader and the task team members' collaboration with the recipient countries' implementation team. Previous literature has emphasized the importance of team collaboration and the quality of the task team leader (Denizer, Kaufmann, & Kraay, 2013; Diallo & Thuillier, 2005; Khang & Moe, 2008; Ika, Diallo, & Thuillier, 2010). Better collaboration within the team members and with a recipient country's government might influence the implementation and also positively affect the project outcomes. Since the information from the quantitative data is very limited, I use qualitative data and methods

to overcome this limitation and observe what other factors largely influence project outcomes.

There is also a concern for possible sample selection bias. First, the sample is limited to those projects that were implemented properly, have been evaluated and received ratings from the IEG. The sample excludes projects that are cancelled and have never received evaluation ratings. Despite many possible reasons for the cancellation of a project, it may be more likely that weakly designed projects or countries with weak implementation capacities have higher probabilities of receiving lower ratings if they had been completed. This may lead to bias in the coefficients of an explanatory variable (e.g. upward bias in the coefficient of government effectiveness). Although random sampling may be a possible solution, it is difficult to do random sampling in my study, as the number of education projects is already small.

Another type of bias due to sampling is raised by previous studies. The sample of projects supported by the World Bank is not random nor representative of all education projects in any borrower country (Isham & Kaufmann, 1999). It is possible that the World Bank or the borrower government skims and identifies the possible projects (Isham, Kaufmann, & Pritchett, 1997; Isham & Kaufmann, 1999). The bias will differ depending on whether the skimming occurs from the government's side or the Bank's side. It is possible that a government selects a project that is most problematic to implement and asks for World Bank financing, and finances the best projects out of their own budget (Isham, Kaufmann, & Pritchett, 1997). In this case, the projects that are selected may be systematically different and tend to have lower ratings compared to the projects that are not selected.

On the other hand, if the World Bank identifies and selects the best possible projects that have high expected rates of return, then the projects in the sample will have higher ratings than projects that are not selected, and have upward bias in any explanatory variable that are expected to have positive relationship with project outcomes (Isham & Kaufmann, 1999). Using the information on the expected rates of return and reestimated rates of return, Isham and Kaufmann (1999) tested for the potential misspecification due to sample selection bias and concluded it does not bias the coefficients of their interest variable. Conducting tests for sample selection bias for this study is not possible as there is no data on expected ratings and actual ratings available to researchers outside the Bank. However, the preliminary interviews with the Bank staff implied that the Bank's internal culture does not place a lot of emphasis on the project ratings and there are no incentives or punishment for staff based on project outcomes. Thus, I regard the threat on the validity of this study is less from the bias occurring from Bank's selection on best projects.

Lastly, the measurement error with my dependent variable can also be a threat for the validity of my finding. This is related to the validity of IEG ratings data. Since the IEG evaluation ratings are not purely conducted by an external evaluator outside the Bank, there can be doubts on whether these ratings reflect the true impact of the project. Since there is no data available on the true impact that can be compared across projects, I use alternative data to test for measurement error in the project outcomes.

The assumption with the measurement error in the dependent variable is that when it is not systematically correlated with any of the explanatory variables, the OLS estimators are consistent and possibly unbiased (Wooldridge, 2010). Thus, I investigate

whether the measurement error in the project outcomes is correlated with any of the explanatory variables used in my model. Since there is no observed data on true project outcomes, I assume that the initial ratings given by the evaluator within the Bank management is my outcome variable. I also assume that the final evaluation rating given by the IEG is the true project outcome.

I randomly selected 30 education projects from the data and coded their ICR, the initial evaluation ratings. Then I compared the ICR ratings to that of the final IEG ratings to see if there is any tendency to over-evaluate or under-evaluate the project outcomes. Table 5[22] below suggests that there is some tendency that the ICR ratings are downgraded by the IEG. In order to explore if the difference is statistically independent from any of the explanatory variables, I ran several regressions using some of the explanatory variables used in my model on the difference of ICR and IEG ratings as the dependent variable (coded as binary, difference=1, no difference=0). I did not find any explanatory variable significantly associated with the difference in the two ratings. Although the sample is extremely small (only 30 projects), I did not find any solid evidence that there is measurement error in the project outcome variable. However, further studies need to address this issue with a larger sample if data is available.

[22] For some reason, there were two projects that did not have ICR rating but only had IEG rating. Therefore, the ICR rating adds up to 28, and IEG rating adds up to 30 projects.

Table 5 Differences between ICR and IEG ratings from randomly selected sample

	Initial outcome (ICR rating)	Final outcome (IEG rating)
Highly satisfactory	2	1
Satisfactory	22	12
Moderately satisfactory	1	8
Unsatisfactory	3	9
Highly unsatisfactory	0	0

3. Research Question 2: Perceived determinants and outcomes

The purpose of the second research question is to examine whether there are other project characteristics that are not included in the IEG ratings data, and to examine the extent to which the outcome might be rated subjectively. As this is considered as a "black box," I apply a naturalistic inquiry to answer this research question. A qualitative study is exploratory and based on the assumption that the concepts are not identified yet (Corbin & Strauss, 2008). Thus, my question is broad: "How do the Bank staff perceive the project determinants and outcomes?"

Sampling

I used a nonprobability sampling method, which is often used in qualitative research. The information on the World Bank staff are not publicly known, although the project documents do mention the task team member's names. It is more difficult for an outsider to contact retired staff or staff that left the Bank. The staff, however, is interconnected if they worked at the Bank around the same period. Therefore, I initially intended to use snowball sampling, which is useful for hard-to-reach or hard-to-identify

populations, but the members are somewhat interconnected (Schutt, 2009). However, given the time constraints, availability sampling was adopted. Interviewees were contacted via email asking for their interest in participating in the study.

I attempted to balance the characteristics of the interviewees in terms of years of experience, their specific position within the education sector, and the region they were mainly in charge of. The majority of interviewees had at least 10 years of working experience at the education sector in World Bank. The estimated average years of experience is 16.4 years. Positions of the interviewees varied, including operational staff, sector manager, regional office staff, and research officer. Nine of the 15 interviewees were currently or used to work as operational staff. Four interviewees were or currently working for the IEG (former OED). The period of when they worked also varied; however, 12 out of 15 interviewees were working between 1985 and 2011, which matches with the evaluation ratings data used in the quantitative part of this study.

Data collection

I conducted semi-structured individual interviews of the current and retired staff of the education sector of the World Bank. Before these interviews were conducted, the interviewees were contacted through email and were informed of the topic of this research, and that their participation in this study is voluntary and the audio-recorded file as well as the transcripts will all be kept confidential. Total 15 staff was interviewed. An initial round of interviews were conducted during March 10–15, 2014. The second round of interviews were conducted at or near World Bank Headquarters in Washington D.C. during May 1–7, 2014. Although most of the interviews were conducted face-to-face, six

interviews were conducted using Skype due to schedule and location constraints. Each interview lasted for 30 minutes to two hours, and was audio-recorded with permission from each interviewee.

A list of initial interview questions is attached in the appendix. Based on the initial questions, I probed and asked further questions based on the interviewee's responses. As Rubin and Rubin (Rubin & Rubin, 2012) suggest, I prepared a limited number of main questions in advance, which are orienting tour questions to get general orientations on the education project cycle, evaluation and project characteristics. I asked follow up questions and probe into an emerging topic or theme. I conducted responsive interviews, which is a style of qualitative interviewing (Rubin & Rubin, 2012). Since the experience of each staff member varied based on the period and region they worked at the World Bank, the set of interview questions differed between interviewees. According to Rubin & Rubin (Rubin & Rubin, 2012), a separate set of questions can be asked to different interviewees since new questions are designed to tap the experience and knowledge of the interviewee. The pattern of questioning is flexible, questions evolve in response to what the interviewees have just said (Rubin & Rubin, 2012, p. 36).

Analytic strategy

For the analysis of this study, I use a constructivist grounded theory approach, an inductive approach in which initial categories are derived from the data, rather than using categories based on theories (Strauss & Corbin, 1990). Concepts and themes emerge from the data without reference to the literature, and they emerge bit by bit (Rubin & Rubin, 2012). Since one of the purposes of my second research question is to compare how the

perceived determinants of project outcomes are different from those observed in the secondary dataset, grounded theory approach is appropriate, as I create new categories based on the data not from the literature. Thus, I will be able to describe the extent to which the perceived project determinants align with the characteristics in quantitative data.

I transcribed all of the audio-recorded interview files in full and summarized each interview shortly after the completion of each interview, as suggested by Rubin and Rubin (Rubin & Rubin, 2012). A full transcript, rather than relying on memory, should not bias my results. I conducted initial coding without considering the theoretical framework suggested by Middleton (1985) to ensure that I have emerging themes derived from the data that I have collected. This inductive approach allowed me to recognize what is happening in the data and to develop an emergent theory to explain the data (Charmaz, 2006). It also allowed me to create categories that are based on the data, instead of making categories based on theories and being forced to fit the data to the categories. The initial coding also allowed me to take a look at the data and get a sense of the whole picture of what the data describes. I did word-by-word coding and as advised by Charmaz (2006), I moved quickly through the data and coded using Nvivo.

Transcripts are coded word-by-word and categories of broad concepts, themes and events were created. Nvivo software was used to code the transcripts. Coding categories were labeled. Coded words were sorted within a larger domain, creating sub domains. The axial coding was conducted in excel spreadsheet. For some categories that contained a broader concept, I created sub-categories based on the data. For example, for the bigger theme category "Quality of Bank staff," I created sub-categories of "ability to

manage resource and team", "Bank staff being on the ground", "cultural tolerance," "experience," "personality," and "understanding education." Relationships between the sub domains and the domains are defined. Then I systematically examined different interviews to clarify concepts and themes. Based on a set of related concepts, I examined whether the relationship is causes, conditions, opposites or consequences (Rubin & Rubin, 2012). The newly created domains (or categories) emerged from the data are integrated with Middleton's (1985) framework.

I also conducted constant comparative analysis to ensure that the same incident falls under the same node. I also added definition of each category (node). I compared my categories/nodes with that of Middleton's (1985) analytical framework and revised his analytical framework based on my data. I added new categories that did not exist in Middleton's analytical framework.

Trustworthiness

Lincoln and Guba (1985) list four categories for trustworthiness of the findings from qualitative data. First, to ensure credibility, I use triangulation and negative case analysis[23]. For triangulation, I used different sources (different interviewees) to confirm same information; however, I gave more weight to the interviewees with more experience, as they have dealt with many education projects. I also gave more weight to those who are currently working as operational staff. In addition, since the time period of the interview data needs to match the IEG ratings data used in the quantitative analysis, I

[23] It would be ideal if member checks could be done, however, due to time constraints; member checks were not feasible to be done.

gave more weight to the projects that were implemented after the 1990s when interpreting the interview data[24].

I also used negative case analysis and continuously refined a hypothesis until it accounts for all known cases without exception (Lincoln & Guba, 1985). Combining these methods should ensure the credibility of the findings.

For transferability, I illustrate a detailed, thick description of interview context and settings. For dependability and confirmability, I kept a reflexive journal throughout the coding process. The reflexive journal is also part of the audit trail. Providing audit trail is helpful to ensure the trustworthiness. For the audit trail, I keep a record of four out of six categories suggested by Lincoln and Guba (1985, pp. 319–320): 1) Raw data (interview guides, notes, recorded file) 2) data reduction and analysis products (summary notes) 3) data reconstruction and synthesis products (grounded theory and data analysis sheets, reports) and 4) process notes (reflexive journal). Audit trail allows an auditor to determine the trustworthiness of my study, thus it should lead to dependability and confirmability (Erlandson, Harris, Skipper, & Allen, 1993).

Potential threats to validity

Social desirability might influence the validity of findings. It is possible that interviewees might provide self-justifying or socially acceptable answers rather than revealing what happened and their role in the events (Lyman, 1998). However, the World

[24] I use the rationale explained by Rubin and Rubin (2012) for weighing the interviews by the senior people. Rubin and Rubin explains that it is possible that one is better witness than the other and in this case researcher will want to rely more heavily on that witness's report who seems to remember more clearly about the event. Considering that a World Bank project is implemented for several years, it is highly possible that senior people have better insights in determining success factors for project outcomes due to their long experience in project operations. It is also possible that senior people are more outspoken on evaluation issues of project ratings than the younger staff. Therefore, I put slightly more weight on the interviews on senior people when interpreting the data.

Bank staff tends to be critical about their own system, as there were internal criticisms on the lending policy in the past. Thus, I am less concerned about the bias in the results due to social desirability.

CHAPTER VI

FINDINGS

Research Question 1: What factors explain the variation of education project performance implemented by the World Bank?

The significance of country characteristics

The overall results show that country level characteristics are significant predictors for probabilities to receive satisfactory ratings than project level characteristics. Government effectiveness, East Asia Pacific region dummy, education spending, primary school enrollment rate and GDP growth as well as some interaction terms with country level variables included turned out to be significant, whereas very few project level characteristics show significant results. However, except for education spending variable, the significance of country level variables are sensitive to model specifications.

Government effectiveness

My first research question is an exploratory question and the results from estimations from OLS, logit (Table 9, 10) show that the government effectiveness variable was a statistically significant predictor for a project to receive satisfactory ratings. The logit estimation (Table 10, column 2) shows that a unit increase in government effectiveness is associated with an average of 0.9 increases in the log odds of

receiving satisfactory ratings, holding all other variables constant. The effect is also practically significant as the odds ratio indicates that for a unit increase in government effectiveness, the odds of receiving satisfactory ratings is 2.46 times larger, holding all other variables constant (Table 10, column 2_OR). The OLS estimation (Table 9, column 2) also shows statistically significant effect of government effectiveness on project outcomes. This indicates that the effectiveness and the quality of governments' delivery of the public services to their citizens significantly matter in project success. However, the effect is not significant in ordered logit estimations (Table 11), which indicates that government effectiveness is sensitive to model specifications. Moreover, the significant effect of government effectiveness goes away when country heterogeneity is taken into account in the models with country fixed effects (Table 12), which is intuitive and expected, as the government effectiveness varies a lot across countries rather than within countries.

Primary school enrollment

Table 9 and 10 shows that primary school gross enrollment is negatively associated with the probability of receiving satisfactory ratings. The logit model result shows that, for one percentage point increase in primary school gross enrollment rate, the log odds of receiving satisfactory ratings decreases by 0.02 on average, holding other variables constant. This finding might seem odd and less intuitive, as greater primary school enrollment should mean better condition for education and thus should be related to better project outcomes. Nevertheless, the rapid expansion of school enrollment may have deteriorated in both educational quality and learning outcomes if the effort to

increase school enrollment did not combine efforts to promote better educational quality. For example, the Bank pressured the government of Mali to introduce double-shift teaching, which permitted rapid increases in enrollment but decreased instructional time per student, which was a major factor in student achievement (Nielsen, 2006). This finding could also be explained by the fact that the countries that achieved rapid increase in enrollment were preponderantly the Sub Saharan African countries, and that 28 percent of the projects in the sample were implemented in SSA region.

One caution for the interpretation of the negative relationship between school enrollment and project outcomes is that the school enrollment variable could potentially be endogenous. Because low-income countries prioritize increasing enrollments and reducing gender inequity in access to schools (World Bank IEG, 2011), it is possible that unobserved country characteristics of low-income countries might have affected negatively to project outcomes. In such case, it is not the primary school gross enrollment rate but the unobserved country characteristics that are the reasons for receiving low project ratings. When the country fixed effects model eliminated the unobserved country characteristics, the significant effect of primary school gross enrollment goes away.

GDP growth

The logit results from Table 10 shows that GDP growth is positively associated with project outcomes. For an annual increase of one percentage point growth in GDP, the log odds of receiving satisfactory ratings increases by 0.11 (the odds is 1.12 times larger on average), holding other variables constant. This result is intuitive, as countries that experience growth in economic development have better environment for

implementing education projects. It is also possible that the governments of those countries going through their economic developments tend to have stronger commitment in leading to successful implementation of projects.

Education spending

Education spending (as % of GDP) is negatively associated with the probability of receiving satisfactory ratings. The results of logistic regression (Table 10, column 2, 2_OR) indicates that a one percentage point increase in education spending as percent of GDP was associated with an average of 0.77 times smaller odds of receiving a satisfactory rating, holding other variables constant. The linear probability model results shows that one percentage point increase in education spending as percent of GDP is associated with 0.04 decrease in the probability to receive satisfactory ratings, holding other variables constant. The result indicates that more spending in education is related to less successful project outcomes, which is not intuitive and against our common sense.

In fact, the education spending variable is likely to be endogenous and that other country level characteristics might be associated with both the education spending and the project outcomes. It is possible that the countries that spend more on education than other countries tend to have a political, economic and social environment that are difficult to implement projects successfully and receive unsatisfactory ratings. This depends on how education spending is reported from each country. If poorer countries (that have less favorable environment for implementing education projects) have smaller GDP than other countries, but receive foreign aid in education and add the amount of aid to their education budget, than it is likely to show that the more countries spend in education as

percent of GDP, the less likely for the education projects to receive high ratings because of the external characteristics that are related to the implementation of education projects. The fixed effects estimation in Table 12 shows that the significant effect of education spending is gone, except for column 3 (logit model).

The full model with interaction terms included (Table 10, column 1) shows that, the interaction term of education spending and primary education is significant throughout all estimates, indicating that there is a differential effect of education spending in primary education sector and non-primary education sector. Education spending for primary education is associated with 1.38 decrease in the log odds of receiving satisfactory ratings, holding other variables constant (education spending [-0.48] + primary [-1.29] +education spending*primary education [0.39] = b = -1.38). The interaction term of education spending and primary education remained significant in logit model (Table 10, column 3) even after country heterogeneity is taken into account. This finding indicates that education spending in primary education projects is associated with 1.75 decrease (Education spending [-0.75] + Primary education [-1.57] +interaction [0.57] =b= -1.75) in the log odds a project being rated as satisfactory, holding other variables constant. This finding indicates that education spending is particularly linked to lower ratings compared to the projects for primary education sector. There might be some project level factors that vary within countries affecting education spending and primary education.

Regional dummies

Results show that East Asia Pacific region is positive and significantly associated with the probability to receive satisfactory ratings. The full logit model from Table 10 indicates that a project implemented in the East Asia and Pacific region has, on average, 6.35 times larger odds of receiving satisfactory ratings than other regions, holding other variables constant. In the ordinal logit full model, the odds were 4.22 times larger than other region, ceteris paribus. These results indicate that education projects in East Asia Pacific region is significantly more likely to receive satisfactory ratings. Similar result was found by Petrosino et al. (2012) who systematically reviewed the randomized controlled trials and quasi-experimental studies that aimed to increase school enrolment. Petrosino et al. (2012) found that interventions conducted in East Asia and the Pacific were associated with larger average effects compared to interventions in other regions. Part of the reason could be associated with cultural factors specifically related to East Asia, such as valuing education among people. Other reasons could be strong government commitment that will to lead project implementation. This finding was expected and aligns with the findings from interview data, thus will be discussed more in the discussion section.

The interaction term of East Asia Pacific region and government effectiveness is also significant in ordered logit estimation results. Table 11 column 3 shows that, for a unit increase in the government effectiveness measure in the East Asia Pacific region, the log odds of receiving a satisfactory rating increases by 3.36 (East Asia Pacific [1.44] + Government effectiveness [0.47]+East Asia Pacific*Government Effectiveness

[1.45]=b=3.36), ceteris paribus. This means that the effect of government effectiveness significantly differs between East Asia Pacific and other regions.

However, an unexpected finding from regional dummies is that Africa dummy is negative but not statistically significant. This is different from the findings from interviews and previous literature (e.g. Kilby, 2012) in which the authors found a significantly negative effect of African dummies on project performances across all sectors. However, the non-significant finding of African dummy aligns with the findings from Dollar and Levin's (2005) study where their sub-sample analysis on education sector also showed non-significant effect of Africa dummy on project success rate. The non-significant finding of the Africa dummy could be because when looking at Figure 11, Project ratings by region, although relatively more education projects implemented in Africa received unsatisfactory ratings (38.17%), still more than half received satisfactory ratings (61.83%). This may indicate that although the environment is difficult to implement projects in Sub Saharan Africa countries, education projects are relatively showing fine results compared to projects in other sectors.

Generally non-significant and mixed findings of project level characteristics

The project level characteristics such as project duration, cost, type of loans, and subsectors are not statistically significant. The effect of project duration is not statistically significant for predicting project performance in models without country fixed effects, meaning that the longer project implementation period is not necessarily associated with better or worse outcomes. The country fixed effects estimations (Table 12, column 1, 2, 4) shows that in the OLS and logit model without interaction terms, project length is

positive and significantly associated with project outcomes, indicating that the longer a project is implemented, the better the project outcomes will be, when country heterogeneity is eliminated and when only country variation is considered. This is a contradictory finding from that of Denizer et al.'s (2013) study, where the authors used IEG data from all sectors and found significantly negative relationship between project duration and project performance. The positive impact of project duration also does not match with the findings from qualitative data, that indicated longer implementation period usually means there is delay in the process, or there have been problems and therefore the project is restructured. Due to mismatch between quantitative and qualitative findings, this study remains inconclusive on the relationship between project length and outcomes. Further exploration is needed regarding the relationship between project duration and outcomes.

Project cost is also not significant throughout all model specifications, implying that there is no evidence that more financial investment in the project is a solution to increase the probability of receiving satisfactory ratings. Project cost is positively associated with project outcomes in OLS and logit estimations, and negatively associated with project outcomes in ordinal logit estimations, but insignificant. This is, again, a different finding from that of Denizer et al.'s (2013) study, where they found the log of loan size was significant and negatively associated with project performance. It is also a different finding from Kilby's (2012) study where he found significant and positive effect of project size (cost) on project outcomes. However, my finding from ordinal logit estimation aligns with that of Dreher et al.'s (2010) study where the project size was negative but not significant.

The non-significant findings of these variables (project duration and cost) may be counter-intuitive as we expect that greater resources and time spent on projects may lead to better results. However, the negative or no partial effect of project length and costs imply that unobserved project characteristics could also affect the project outcomes (Denizer, Kaufmann, & Kraay, 2013). For example, relatively more "ambitious" projects are much more difficult to implement, and therefore, take longer time and money to complete, but also may be more likely to turn out to be unsatisfactory, if not "no effect" on project outcomes (Denizer, Kaufmann, & Kraay, 2013). Thus, the project duration and cost is likely to be endogenous.

The dummy for development policy loan is negative but not significant (Table 9, 10), indicating that there is no significant difference in project outcomes between development policy loan and specific investment loan. This finding partly aligns with the findings from Denizer et al.'s (2013) study, where the authors found no significant relationship between dummy for investment projects (specific investment loans) and project outcomes of projects from 1995–2011. However, they found a significant and positive relationship between dummy for investment projects and project outcomes with binary outcome ratings of projects from 1983–2011.

The difference may be coming from the nature of education projects and non-education projects. Education projects tend to be successful when they are combined with other efforts. An education project that aims to increase school enrollment in the rural area will be successful not only because of the high quality of the project but also support from the government in other subsectors such as raising teacher quality, providing textbooks and learning materials will be needed. This requires the nature of education

project to be implemented as a sector wide project, although one is defined as specific investment project. Thus, the sector-wide characteristic of education projects could be related to the non-significant results of the difference between development policy loans and specific investment loans.

The dummy for IDA is also not statistically significant, implying there is no evidence that there is significant difference in education project performance between IDA and IBRD funded projects. Since IDA provides loans to the least developed countries, it is possible to expect that projects tend to show greater influence in the poorest countries. However, this finding is unexpected, and also contradictory to the finding of Kilby's (2012) study, which found a positive and significant relationship between IDA countries and project performances. One reason could be related to the non-significant findings of African and South Asian regional dummies. Most of the IDA loans are provided to Africa and South Asia (Gilbert & Vines, 2000), but the regional dummies did not show any significant relationship with the project outcomes, therefore, the IDA dummy does not show significant findings. Another reason could be that IDA receiving countries are much more vulnerable to political and economic stability while they have more potential to receive larger effect of a project in the community. If these two factors cancel out, it is possible to find that IDA dummy is not significant.

Repeater projects, referring to the projects that are in the second or third phase of an initial project, are also not significant, which aligns with the findings of Denizer et al. (2013). This is also counter-intuitive, as it may seem natural that a project that builds upon a former one would have better results. However, this finding, again, may be related to the unobserved country and project level characteristics. Projects that are difficult to

implement may tend to receive lower ratings regardless of whether they are in the second or third phase. Projects that are being implemented in fragile states would get lower ratings regardless of whether the project is a repeater.

Education sector dummy is also not significant, as Kilby's (2012) study also found that dummy for "other sector" was not significantly associated with project outcomes. However, this finding does not align with Kilby's (2000) earlier study, where he used the project performance rating in the 1980s and found a positive effect of education sector on change of project performance. One of the reasons could be related to the decline in project ratings since the 1990s and 2000s, when a large number of poor-performing Learning and Innovation Loans existed (World Bank IEG, 2011).

The share of the largest subsector is negative but not significant in all model specifications. Denizer et al. (2013) found a negative and significant relationship of share of projects in the largest sector and project outcomes. One of the reasons for the difference in findings from my study and Denizer et al.'s (2013) study could simply be due to sample size. Denizer et al. (2013) were looking at projects across all sectors, thus had a sample size of 6,569 projects, whereas my study is only looking at education sector and thus includes less than 500 projects. Thus, there is no evidence from my study that there is a relationship between projects that focus only on a specific subsector and project outcomes. However, the findings from the interview data show a different result, as will be discussed in the next section.

The types of loans added in the models (emergency recovery loan, learning and innovation loan, specific investment loan, technical assistance loan and structural adjustment loan) generally showed no significant association with project outcome across

different estimations. With the data from 1986–2011 (Table 8, column 1–6) and in country fixed effect estimations (Table 12 column 1, and 2), the learning and innovation loan showed a significantly negative relationship with project outcomes. This finding was not consistent in Table 9, 10 and 11 with data from 1996–2011 and estimations without country fixed effect. This finding is somewhat expected since the histogram of ratings for learning and innovation loan (Figure 12) shows that half the loans are rated as unsatisfactory while more than 70 percent are rated as satisfactory in other loan type projects. However, this finding may not be robust throughout different specifications as the total number of projects implemented from learning and innovation loan is very small (only 28 projects), and the small sample size might have limitations in detecting statistical significance.

The ordinal logit model specifications show negative and significant relationship between structural adjustment loan and project outcomes. Table 11 column 3 indicates that a structural adjustment loan is associated with an average of 1.07 decrease (0.34 times smaller odds) in the log odds of receiving satisfactory ratings than other type of loans, holding other variables constant. The negative relationship between structural adjustment loan and project performance contradicts Kilby's (2012) study, where he found a positive and significant effect for project outcomes across all sectors. This is because the education sector is disadvantaged from structural adjustment loans that were conditioned on policy changes (Mundy & Verger, Forthcoming). Caps were put in government spending as conditionalities focused on cuts to the government's civil service (Mundy & Verger, Forthcoming). However, the finding of negative and significant

effects of structural adjustment loan is not consistent with binary outcomes of projects (see Table 9 OLS, Table 10 Logit model results).

The dummies for subsectors (primary, secondary, tertiary, general education) are also not significant throughout all estimates. Only the data from 1986–2011 in Table 8 shows a negative and significant effect of primary sector, however, since the model includes an interaction term of education spending and primary, the interpretation of this finding is meaningless. The non-significant finding of subsectors is expected as Figure 13 shows that regardless of project subsectors, most of the projects (over 70 percent) received satisfactory ratings and only about 20 percent of projects received unsatisfactory. The IEG (World Bank IEG, 2011) reported that one of the reasons for downward trend in performance is due to changing portfolios within education sector. However, the non-significant finding of subsector dummies in this study indicates that there is no evidence to conclude that post-primary projects, either secondary, tertiary or vocational education projects are more difficult to implement than primary education projects. In fact, the changing portfolios in education project is a natural outcome as countries reach or at least approach universal primary education and post-primary education became more important to national development goals.

One consistent finding throughout most of the model specifications is that the year 2008 and year 2011 is negative and significantly related to project outcomes. This indicates that there is an external factor associated with the year 2008 and the year 2011 that has negatively affected the education project ratings. The global financial crisis that occurred around 2008 might have directly affected the economies of developing countries, leading to disruptions in operating the projects. The financial crisis has also indirectly

affected the economies of donor countries that decided to de-prioritize education within their aid budgets, which led to a decline in the share of education in total aid (UNESCO, 2013a). Within the decline of aid in education, total aid for primary education declined more, and the decline was particularly strong for Sub Saharan African countries (Steer & Smith, 2015). Since the poorest countries tend to rely heavily on external financial aid, the reduced aid to education after 2010 may have been a factor for the negative and significant relationship between the year 2011 and project outcomes.

Predicted probabilities

In order to look at the value of the dependent variable (project ratings) with given values of independent variables, I look at the predicted probabilities. As explained before, since the logit is the natural log of odds, the interpretation of coefficients from logistic regressions are not intuitive. Therefore, it is necessary to look at the results that are transformed back to probability scale, which is the predicted result of logistic regression (Peng & So, 2002). Predicted probabilities can be computed at given values of independent variables[25] (Long & Freese, 2006). Figure 14 shows the predicted probabilities for each observation ranges from 0 to 1 but that most observations have predicted probabilities between 0.6 and 1.

Since it is more effective to calculate predicted probabilities at given values of independent variables, I show the predicted probabilities of receiving satisfactory ratings by specific project types. Table 6 illustrates that a project with all independent variables at its mean value has probability of receiving satisfactory ratings between 0.75 and 0.84.

[25] Since predictions are computed for all cases that do not have missing values, the predicted probabilities are calculated with the data set that is not imputed for missing values.

It also shows that a secondary education project implemented with a specific investment loan in a IDA country has 0.76 probability of receiving satisfactory rating. The table also shows that an emergency loan and a project that is in its second phase or beyond has higher probabilities than a non-emergency loan or a project that is implemented for the first time.

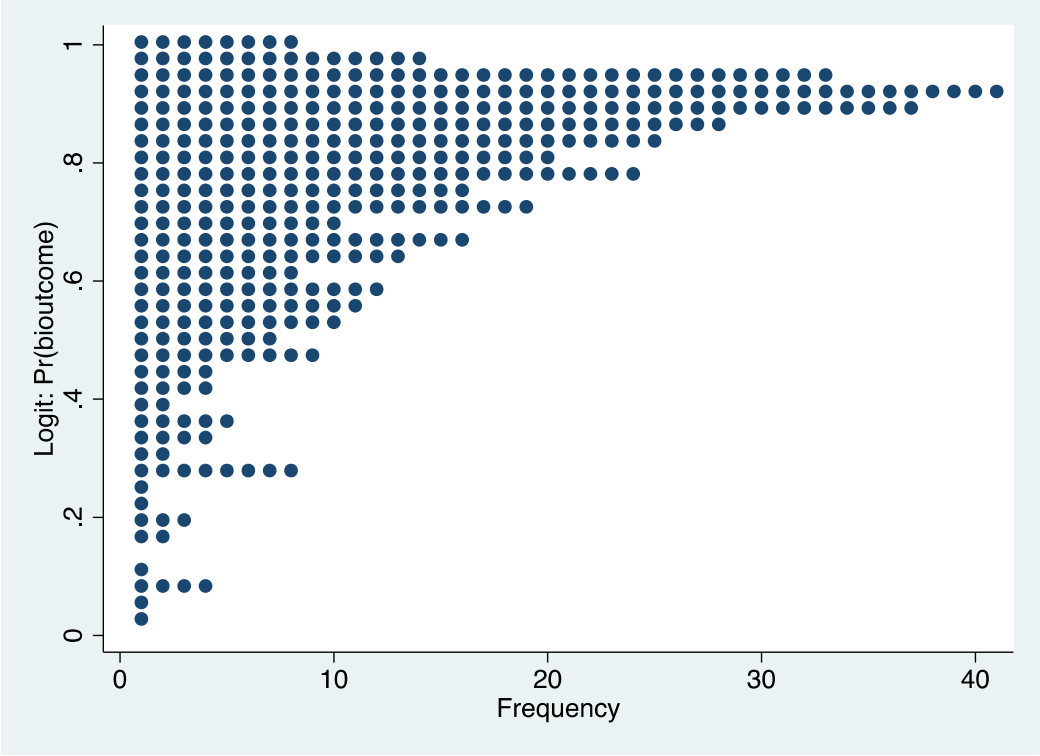


Figure 14 Predicted probabilities for individual observations

Table 6 Predicted probabilities of receiving satisfactory ratings by specific projects

Type of projects	Probability of receiving satisfactory ratings (95% CI)
An "average" project	0.8 (0.75, 0.84)
Projects implemented in IDA countries, is a specific investment loan, aimed for primary education sector	0.63 (0.35, 0.91)
Projects implemented in IDA countries, is a specific investment loan, aimed for secondary education sector	0.76 (0.54, 0.98)
Projects implemented in IDA countries, is a specific investment loan, aimed for tertiary education sector	0.73 (0.51, 0.95)
Projects implemented in IDA countries, is a specific investment loan, aimed for vocational education sector	0.8 (0.52, 1)
A repeater project in primary education	0.75 (0.51, 0.99)
A non-repeater project in primary education	0.61 (0.36, 0.87)
A repeater project in secondary education	0.85 (0.73, 0.96)
A non-repeater project in secondary education	0.76 (0.62, 0.9)
A repeater project in tertiary education	0.82 (0.68, 0.96)
A non-repeater project in tertiary education	0.73 (0.56, 0.89)
An emergency recovery loan (ERL) in primary education	0.87 (0.64, 1)
A non-ERL in primary education	0.64 (0.38, 0.89)
An emergency recovery loan in secondary education	0.92 (0.79, 1)
A non-ERL in secondary education	0.77 (0.64, 0.91)
An emergency recovery loan in tertiary education	0.91 (0.75, 1)
A non-ERL in tertiary education	0.74 (0.58, 0.9)
A project implemented when a country has low primary school enrollment, low education spending as percentage of GDP, and low level of government effectiveness	0.71 (0.52, 0.89)
A project implemented when a country has high primary school enrollment, high education spending as percentage of GDP, and high level of government effectiveness	0.86 (0.76, 0.96)

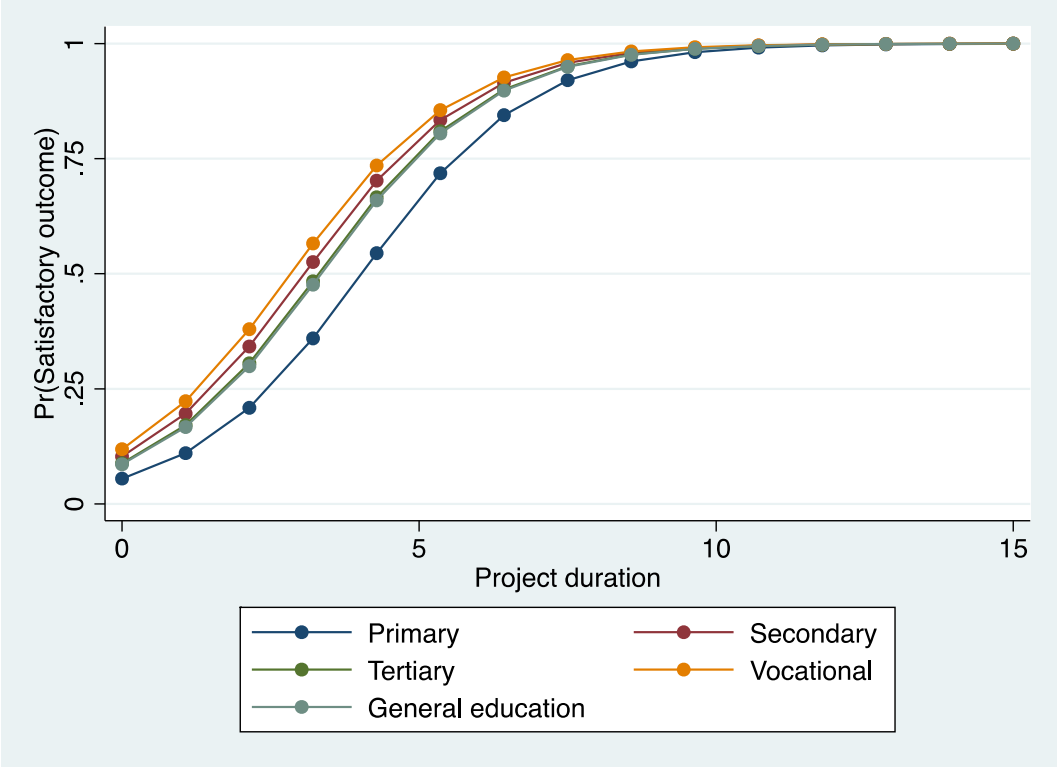


Figure 15 Predicted probabilities as project duration changes by subsector

Figure 15 shows that as the duration of projects are longer, particularly between 0 and 5 years, the probability of receiving satisfactory ratings increases; however, this relationship does not differ much across subsectors. On the other hand, Figure 16 below shows that the change in the probabilities as the duration of a project gets longer is largely different across levels of government effectiveness. The projects with low government effectiveness have much room for improvement as the duration increases compared to the projects with high government effectiveness.

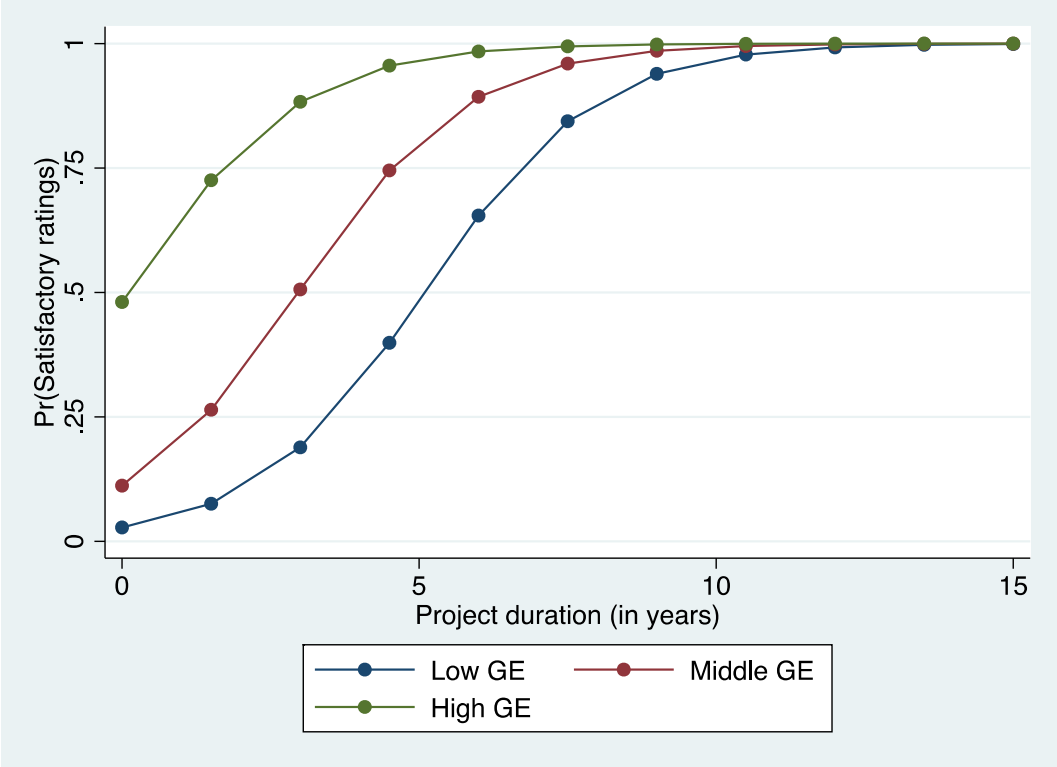


Figure 16 Predicted probabilities of receiving satisfactory ratings as project duration changes by level of government effectiveness

Figure 17 shows that as the share of the largest subsector increases, the probability of getting satisfactory ratings decreases, but it differs by the level of government effectiveness. Countries with high levels of government effectiveness tend to decrease less than the countries with low level of government effectiveness. This suggests that the projects with small share of the largest sector (projects that are implemented across multiple sectors and are less focused) tend to perform better than the projects with a large share of a specific sector (projects that are focused on a subsector). However, in a country with high government effectiveness, the difference of project ratings between small and large share of a subsector is smaller than countries with low level of government effectiveness. This implies that the share of the largest subsector of a project would not matter so much if a country's government effectiveness is high.

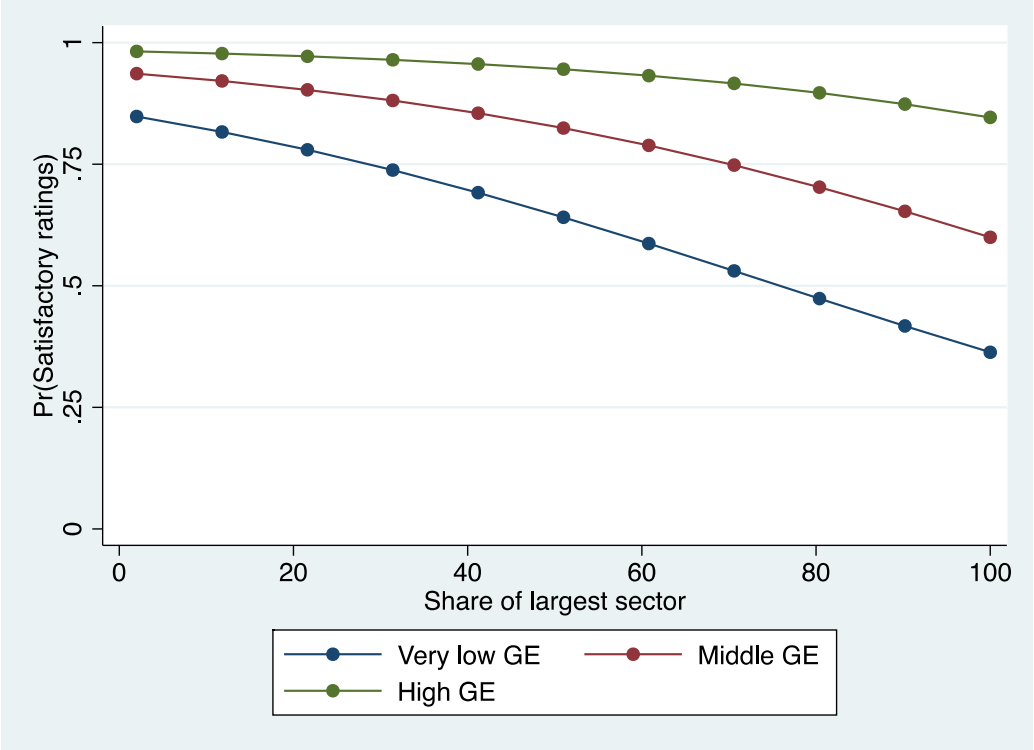


Figure 17 Share of largest subsector and project performance by level of government effectiveness

The predicted probabilities of education spending on probability to receive satisfactory ratings between primary education and non-primary education projects (Figure 18) indicates that primary education sector projects decrease faster than that of non-primary education sector projects as the education spending increases. This suggests that primary education sector projects are more easily influenced by the amount in education spending compared to other sector projects.

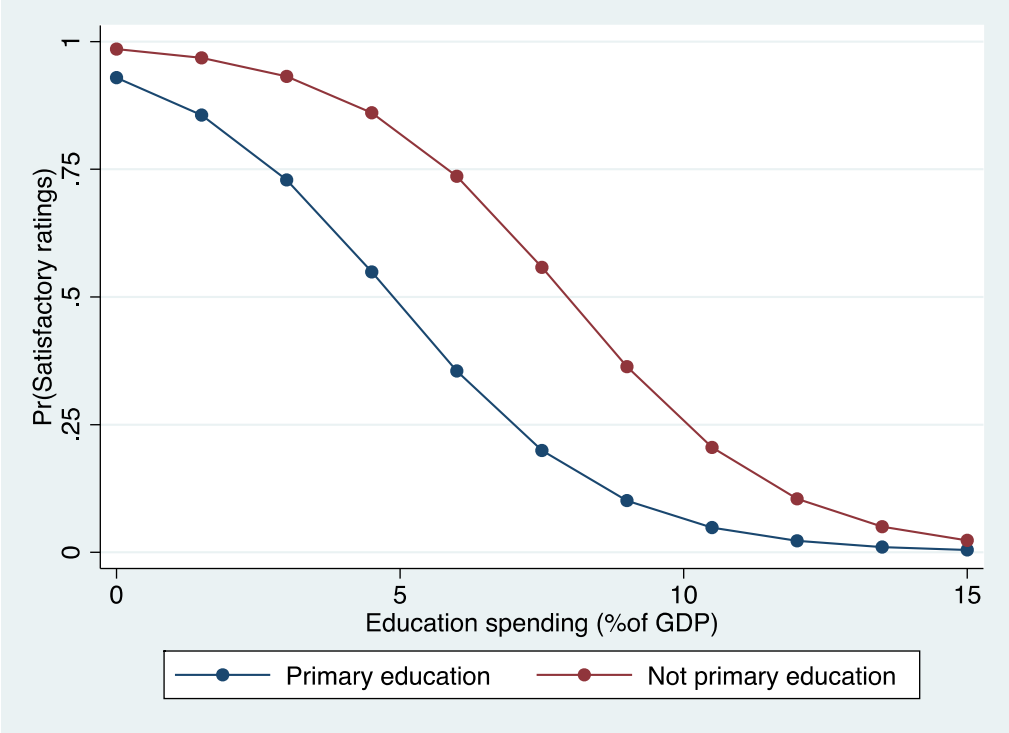


Figure 18 Education spending and project performance of primary education and non-primary education projects

Earlier in the results of ordered logistic regression in Table 11, I found that education spending was negatively associated with the probability of receiving satisfactory ratings. Since the interaction term of East Asia Pacific regional dummy variable and the government effectiveness was positively significant, I explored the predicted probabilities of receiving satisfactory ratings as education spending increased. More specifically, I explored how the decrease of probability of satisfactory ratings differed between East Asia and non-East Asia regions as well as how the probability differs across the level of government effectiveness.

Figure 19 shows that the country located in the East Asia Pacific region with high government effectiveness tends to have higher performance ratings as education spending increases compared to the countries located in regions other than the East Asian Pacific region with low level of government effectiveness. The graph supports that the

decreasing effect of education spending is much less for countries that are in the East Asia Pacific region with high government effectiveness than the countries not located in the East Asia Pacific region. However, the graph also shows that government effectiveness is a stronger factor than the location of the country in the East Asia Pacific region. The red line (non EAP region with high GE) is above the blue line (EAP region with low GE). This indicates that a country with a high level of government effectiveness in non-EAP region performs better than a low level of government effectiveness in an EAP region. Thus, there is no evidence to conclude that simply being in an EAP region is associated with better ratings, instead the government effectiveness was a decisive factor in determining whether an education project has high probability to receive satisfactory ratings.

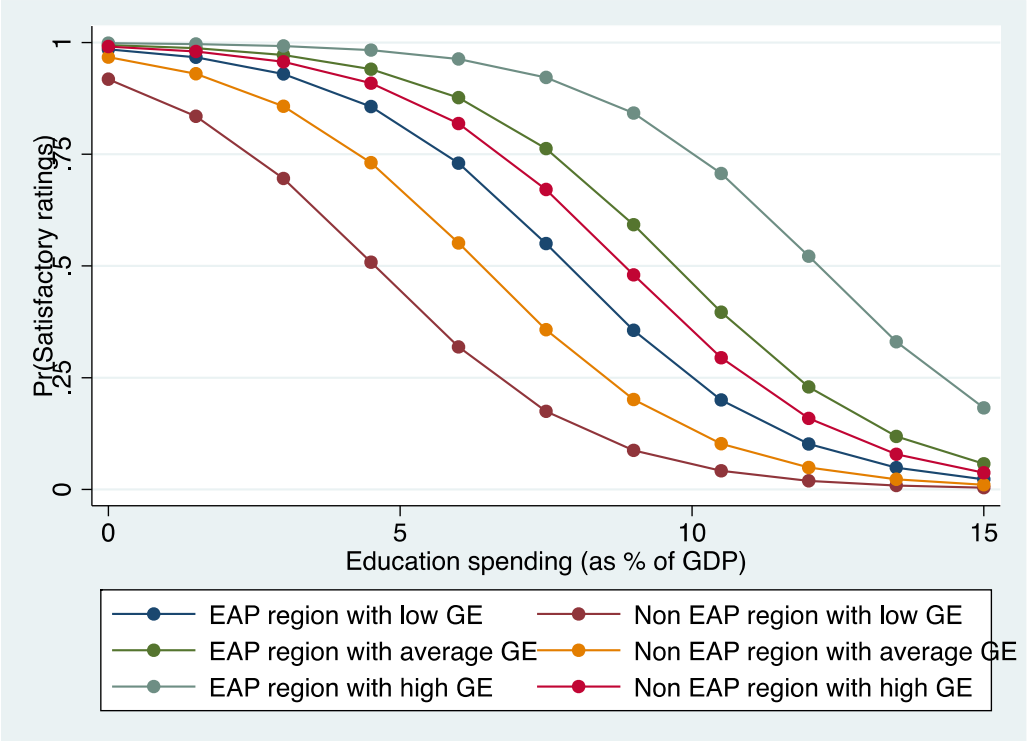


Figure 19 Project performance and education spending between EAP and non-EAP region by level of government effectiveness

For the variables that are not continuous, it is better to look at the marginal effects, i.e. the change in the outcome as one unit in the independent variable changes. For example, interpreting how much the probability changes when a repeater project (repeater=1) becomes a non-repeater project (repeater=0) gives a better idea. Thus, I also present the marginal effects of logistic regressions as Table 7. Table 7 shows the average marginal effects that shows the changes in the probability of receiving satisfactory ratings when all independent variables are held at its mean.

Table 7 shows that an instantaneous change in government effectiveness increases the probability of receiving a satisfactory outcome by 16.7 percentage points, and one percentage point change in education spending decreases the probability by 7.6 percentage points. One percentage point increase in primary school gross enrollment is associated with a 0.43 percentage point increase in the probability of receiving satisfactory ratings. One percent GDP growth increases the probability by 1.4 percentage points. Among the project level variables, Table 7 shows that one year of project duration increases the probability to receive satisfactory ratings by 10.4 percentage points. Also, a repeater project has about 8 percentage point higher probability than a non-repeater project. However, the results are not directly comparable to the results in Table 10 and 11 due to different data set used (imputed data in Table 10 and 11, and non-imputed data for marginal effects).

Table 16 shows the marginal effects for each outcome in ordinal logistic regression models given that all the variables are set at their mean values. The probability to receive satisfactory rating is the highest (37 percent), while the probability to receive unsatisfactory rating is only 5.7 percent. Also, secondary and tertiary education projects

were more likely to receive "satisfactory" ratings (46 percent) than primary education project (27 percent). Primary education projects were more likely to receive "moderately satisfactory ratings."

Table 7 Marginal effects from logistic regression

VARIABLES	(1) Marginal effects
Primary school gross enrollment (logged)	0.432**
	(0.168)
GDP growth	0.0144*
	(0.00840)
Education spending	-0.0759***
	(0.0280)
Government effectiveness	0.167**
	(0.0787)
Project length	0.104*
	(0.0599)
Project cost (USD millions, logged)	0.000704
	(0.0206)
Development Policy Loan	-0.107
	(0.119)
IDA	0.0525
	(0.0971)
Education sector	-0.0780
	(0.0590)
Repeater project	0.0792*
	(0.0420)
Share of largest subsector	-0.0234
	(0.0608)
Primary education	-0.167
	(0.118)
Secondary education	-0.0160
	(0.0688)
Tertiary education	-0.0441
	(0.0718)
Vocational education	0.0118
	(0.105)
General education	-0.0546
	(0.0610)
Observations	404

Note: Standard errors in parenthesis *** p<0.01, ** p<0.05, * p<0.1
Note: Discrete change from the base level
Note: Regional dummies, year dummies and loan type dummies were included in the model but omitted in the table. The data used to compute marginal effects is not imputed for missing data

Subsample analysis

It is necessary to conduct analysis by subsamples, specifically, by subsectors and by loan types. Although Figure 13 illustrates that project ratings do not differ by subsectors, theoretically, it is necessary to conduct analysis by subsectors. It is without any doubt that the characteristics of primary education projects are different from that of tertiary education projects; therefore, they might affect project outcomes differently. For example, the total cost and duration may be more important for tertiary education projects than primary education projects, if both projects aim to improve the infrastructure of schools or universities. Universities are bigger in terms of size, and therefore, tertiary or higher education project would cost much more than primary education projects and it would require a longer period for implementation, which may lead to lower evaluation ratings. Similarly, the characteristics of specific investment loan projects are different from that of sector adjustment loan projects[26].

Table 13 shows the logistic regression results by subsectors (primary, secondary, tertiary, vocational) for projects implemented between 1986–2011[27]. In the primary education sector, project cost showed a positive and significant relationship with project ratings. When the project cost increases by 1 percentage point, the log odds of receiving satisfactory ratings increase by 0.003, holding other variables constant, which is a very small increase. The finding implies that primary education projects with larger costs are significantly more likely to be rated as "satisfactory."

[26] Although the characteristics of projects by loan types can be different, I do not conduct subsample analysis by loan types due to small sample size.

[27] My main results are based on projects approved between 1996 and 2011, however, for subsample analysis I conducted analysis using projects approved between 1986 to 2011, to increase the sample size.

In secondary education sector, an unexpected finding is the positive and significant effect of the MNA region projects, meaning that secondary education project in the MNA region is significantly performing better than secondary projects in other regions. The significant finding of the interaction term of MNA region and education spending indicates that there is a differential effect of secondary education projects in MNA region across the level of education spending as percent of GDP. A secondary education project being implemented in a MNA region in a country that spend more on education as percent of its GDP would have 8.98 (MNA [b=10.19]+Education spending [b=0.22]+MNA*Education spending [-1.43]) higher log odds of receiving satisfactory ratings, holding other variables constant.

The results from tertiary education sector projects show that tertiary education projects are significantly performing better in East Asia Pacific and Latin America and the Caribbean region. A tertiary project, when located in East Asia and Pacific, would have 3.36 higher log odds (or 28.68 higher odds) to receive satisfactory rating, and when implemented in the Latin America and Caribbean, it will have 2.51 higher log odds (or 12.3 higher odds) to receive satisfactory ratings.

Table 13 column 4 shows that none of the observed factors are significant for vocational education projects. More studies need to explore further the factor that contributes most to the satisfactory performance of vocational education. "General education" projects show significantly negative (64.61) odds to receive satisfactory rating when implemented in the MNA region. As it is for secondary education projects, there is a differential effect of general education projects implemented in MNA region across the amount of education spending as percent of GDP. A general education project in an

MNA region in a country that spends more on education spending has 56.4 smaller log odds (MNA [b=-64.61]+education spending [b=-0.46]+ MNA*education spending [b=8.67]) of receiving satisfactory ratings. This finding contradicts secondary education projects where they had positive relationship with education spending in MNA countries.

Sensitivity analysis

To check if the main results change significantly with alternative specifications or other project and country characteristics added in the model, I conducted several other regressions. First, since the government effectiveness turned out to be a positive and strongly associated with education project outcomes in Table 9 and 10 (OLS and logistic regression models), I tested to see if the effect is gone when I 1) exclude the outliers in project performance (in Figure 17, project ratings by government effectiveness), 2) when I include other measures of governance such as political stability, regulatory quality, rule of law, control of corruption, level of democracy, state fragility, and the net amount of ODA the country receives. I ran models without the outliers (the projects that received satisfactory and government effectiveness is very high, above 1), and the government effectiveness variables were still significant. There are about ten outliers at the high end (seven of these projects were implemented in Chile, one in Malaysia, one in Barbados, one in Maldives). I ran the models without these countries, but the government effectiveness variable was still significant.

Figure 20 Project ratings by government effectiveness

I have also examined how the results change by adding other measures of governance. Table 14 shows logistic regression results with additional project and government measures added, and Table 15 shows logistic regression results with additional country characteristics added. The government effectiveness variable stayed significant in most of the model specifications, however, were not significant in Table 14 (5), (7) and (8), when perception on political stability and violence, rule of law and control of corruption variables were added. This is somewhat expected as these variables are strongly correlated with government effectiveness (higher than 0.7). Thus, if both of these variables are added in the model, it is likely that multicollinearity occurs, leading to bias the standard errors. Therefore larger standard errors in Table 12 (5), (7) and (8) may have eliminated the significance of government effectiveness.

Secondly, as it can be seen from the interview results in the next section, government commitment and ownership turned out to be an important factor for perceived project success. I manually coded the proportion of government financing (the actual amount disbursed as opposed to planned commitment amount out of total project cost from the ICR), created a variable "counterpart funding[28]", and used it as a proxy for government ownership, leadership and commitment. However, the results from Table 14 (1) show that the government's counterpart funding are not significantly associated with project outcomes, although the model fit measures improved.

Thirdly, as some of the previous studies (i.e. Kilby, 2000) focused on supervision weeks and found somewhat significant effect of staff time spent on supervision, in order to see if Bank staff's supervision weeks are important in education projects, I coded the proportion of time spent on supervision out of total weeks that Bank staff spent on each project from ICR. I ran logistic regression with the staff supervision weeks added in the model, however, found no evidence that supervision is significantly associated with education project outcomes.

Since the international donor community tends to focus their activities on basic education, a dummy variable for the Education for All theme project was attempted to see if EFA projects performed significantly better or worse than other theme projects. There was no evidence to support the hypothesis that EFA projects performed significantly better than non-EFA projects.

Table 14 shows that government effectiveness is significant when the variable of perceptions on voice and accountability and regulatory quality is included in the model.

[28] The total cost of project includes not only the financing from the World Bank but also the financing from other multilateral and bilateral donors involved in the World Bank project.

This means that even when the citizens' perception on them being able to select their government, freedom of expression and freedom of association and free media (the definition of voice and accountability) is accounted, the government effectiveness is a significant determinant for project performance.

In Table 15, the effect of government effectiveness is gone when the measure of democracy (polity2) and state fragility index is included in the model. The polity index is created by the Center for Systematic Peace and it rates the levels of democracy and autocracy based on general qualities of political institutions and processes (Marshall & Cole, 2011). The index is a combined single measure of regime governance, which ranges from -10 (fully institutionalized autocracy) to 10 (fully institutionalized democracy) (Marshall & Cole, 2011). The finding determines if a country's governance is democratic or autocratic matter for government effectiveness and project performance.

State fragility index is also created by the Center for Systematic Peace and it ranges from 0 to 25, with 25 being "extreme fragility" to signify the level of fragility for a government across eight indicators. A country's fragility is closely associated with its state capacity to manage conflict; make and implement public policy; and deliver essential services and its systemic resilience in maintaining system coherence, cohesion, and quality of life; responding effectively to challenges and crises, and continuing progressive development" (Marshall & Cole, 2011, p. 7). The result (government effectiveness is gone when state fragility index is added) may be because of similar definition of government effectiveness and state fragility index. Although government effectiveness variable captures the perceptions of quality of public services, the definition still overlaps with that of state fragility index. Both indicators are about the government

and state's capacity to formulate and deliver public services. The difference is that state fragility index includes the government's capacity to manage conflict and respond effectively to crises, implying that these can be associated with both government effectiveness and project performance.

Developing countries with larger populations can be a difficult environment for the implementation of education projects, as the quality of education can be low with larger number of students. The effect of population can withdraw the significance of government effectiveness, thus, to test this hypothesis, I added population for each country in the regression model. I found that government effectiveness is still significant even after accounting for the population, and there is no evidence that the population of developing countries is a significant determinant for project outcomes.

It is also possible that the amount of foreign aid that a country receives can be a determinant for project outcomes, or can affect both government effectiveness and project outcomes. A country that is dependent on foreign aid has negative impact on local institutions and government capacity (Berg, 2000; Fredriksen, 2011b); it is necessary to control for the amount of foreign aid, and it can have negative effect on project performance through weakened government capacity. However, Table 15 shows that the government effectiveness is still significant after controlling for the net ODA of recipient countries.

In addition to the variables in Table 14 and 15, I have used alternative measures for GDP growth and education spending as percent of GDP. Since the World Development Indicator provided various measures at the country level, I attempted to use

GDP per capita instead of GDP growth[29], and education spending as a percentage of total government spending in public expenditure instead of using education spending as percent of GDP. I have also included Gini index of inequality in the model. None of the alternative specifications showed any large difference from the main results presented in this study.

[29] Future studies should try using GDP growth per capita, as population growth is important for developing countries, particularly Sub Saharan African countries.

Table 8 OLS, logit, and ordinal logit regression results for all projects approved between 1986and 2011

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Model	OLS	OLS	Logit	Logit_OR	Logit	Logit_OR	Ologit	Ologit_OR	Ologit	Ologit_OR
Dependent variable: Binary project evaluation rating	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	0-5 rating	0-5 rating	0-5 rating	0-5 rating
Project level characteristics										
Project length	0.06	0.07	0.36	1.43	0.41	1.50	0.54	1.72	0.58	1.78
	(0.07)	(0.07)	(0.37)	(0.53)	(0.38)	(0.57)	(0.38)	(0.66)	(0.39)	(0.69)
Project cost (logged, in USD millions)	-0.01	-0.01	-0.04	0.96	-0.04	0.96	-0.10	0.90	-0.11	0.89
	(0.02)	(0.02)	(0.11)	(0.11)	(0.10)	(0.10)	(0.08)	(0.08)	(0.08)	(0.07)
Dummy for Development Policy Loan	-0.02	-0.04	-0.18	0.83	-0.27	0.76	-0.10	0.90	-0.17	0.84
	(0.09)	(0.10)	(0.58)	(0.49)	(0.61)	(0.46)	(0.45)	(0.40)	(0.46)	(0.39)
Dummy for IDA project	-0.05	-0.04	-0.30	0.74	-0.24	0.78	-0.52	0.60	-0.48	0.62
	(0.07)	(0.07)	(0.47)	(0.35)	(0.46)	(0.36)	(0.33)	(0.20)	(0.33)	(0.20)
Dummy for education sector	-0.01	-0.01	0.01	1.01	0.01	1.01	0.15	1.16	0.17	1.19
	(0.05)	(0.05)	(0.29)	(0.29)	(0.29)	(0.29)	(0.25)	(0.29)	(0.24)	(0.29)
Dummy for repeater project	0.01	0.01	0.05	1.05	0.03	1.03	-0.27	0.76	-0.29	0.75

	(0.04)	(0.04)	(0.30)	(0.32)	(0.30)	(0.31)	(0.21)	(0.16)	(0.20)	(0.15)
Share of largest subsector in pct (logged)	-0.04	-0.04	-0.29	0.75	-0.28	0.75	-0.16	0.85	-0.18	0.83
	(0.05)	(0.05)	(0.38)	(0.29)	(0.38)	(0.28)	(0.28)	(0.24)	(0.28)	(0.23)
Emergency Recovery Loan	0.14	0.15	1.1	2.99	1.13	3.08	1.34*	3.81*	1.35*	3.85*
	(0.10)	(0.10)	(0.79)	(2.36)	(0.79)	(2.43)	(0.61)	(2.32)	(0.61)	(2.35)
Learning and Innovation Loan	-0.23*	-0.25*	-1.27	0.28*	-1.31*	0.27*	-0.80	0.45	-0.87	0.42
	(0.11)	(0.10)	(0.58)	(0.16)	(0.57)	(0.15)	(0.46)	(0.21)	(0.45)	(0.19)
Specific Investment Loan	-0.04	-0.05	-0.27	0.76	-0.34	0.71	0.08	1.08	0.03	1.03
	(0.06)	(0.06)	(0.39)	(0.30)	(0.38)	(0.27)	(0.23)	(0.25)	(0.23)	(0.23)
Technical Assistance Loan	-0.15	-0.15	-0.80	0.45	-0.79	0.45	-0.06	0.94	-0.06	0.94
	(0.12)	(0.12)	(0.65)	(0.29)	(0.64)	(0.29)	(0.51)	(0.48)	(0.52)	(0.49)
Structural Adjustment Loan	-0.06	-0.06	-0.36	0.70	-0.34	0.71	-0.87*	0.42*	-0.86*	0.42*
	(0.09)	(0.09)	(0.47)	(0.33)	(0.46)	(0.33)	(0.39)	(0.16)	(0.39)	(0.16)
Primary	-0.14*	0.05	-0.98	0.38	0.32	1.37	-0.60	0.55	0.16	1.18
	(0.08)	(0.05)	(0.62)	(0.23)	(0.29)	(0.39)	(0.39)	(0.21)	(0.21)	(0.25)
Secondary	0.00	-0.00	-0.06	0.94	-0.02	0.98	0.14	1.15	0.15	1.16
	(0.06)	(0.06)	(0.37)	(0.35)	(0.37)	(0.36)	(0.26)	(0.30)	(0.26)	(0.30)
Tertiary	0.01	0.01	0.03	1.03	0.03	1.03	0.34	1.41	0.34	1.41
	(0.06)	(0.06)	(0.36)	(0.37)	(0.35)	(0.36)	(0.28)	(0.39)	(0.28)	(0.39)
Vocational	0.03	0.03	0.18	1.20	0.19	1.21	0.59	1.81	0.59	1.81
	(0.08)	(0.08)	(0.46)	(0.55)	(0.45)	(0.55)	(0.35)	(0.63)	(0.35)	(0.63)

General education	-0.04	-0.03	-0.26	0.77	-0.19	0.83	0.34	1.40	0.37	1.44
	(0.06)	(0.06)	(0.36)	(0.28)	(0.36)	(0.30)	(0.30)	(0.42)	(0.30)	(0.44)
Regional dummies										
Africa	-0.06	-0.06	-0.28	0.75	-0.31	0.73	-0.15	0.86	-0.19	0.83
	(0.10)	(0.10)	(0.56)	(0.42)	(0.55)	(0.40)	(0.40)	(0.35)	(0.40)	(0.33)
East Asia Pacific	0.11	0.13	0.88	2.41	1.00*	2.71*	0.70	2.02	0.80	2.22
	(0.08)	(0.08)	(0.49)	(1.17)	(0.47)	(1.28)	(0.49)	(0.99)	(0.48)	(1.07)
South Asia	0.12	0.12	0.77	2.16	0.8	2.23	0.51	1.67	0.53	1.70
	(0.09)	(0.10)	(0.60)	(1.30)	(0.61)	(1.37)	(0.40)	(0.67)	(0.41)	(0.70)
Latin America and Caribbean	0.11	0.12	0.65	1.92	0.65	1.92	0.48	1.62	0.47	1.59
	(0.09)	(0.10)	(0.51)	(0.98)	(0.52)	(0.99)	(0.42)	(0.67)	(0.42)	(0.67)
Middle East and North Africa	-0.25	0.03	-1.54	0.21	0.17	1.18	-1.63	0.20	-0.18	0.84
	(0.41)	(0.14)	(2.00)	(0.43)	(0.74)	(0.87)	(1.25)	(0.24)	(0.58)	(0.48)
Country level characteristics										
Primary school gross enrollment (logged)	-0.18	-0.19	-1.00	0.37	-1.02	0.36	-0.86	0.42	-0.91	0.40
	(0.12)	(0.11)	(0.64)	(0.23)	(0.60)	(0.22)	(0.50)	(0.21)	(0.48)	(0.19)
GDP growth (annual %)	0.02*	0.02*	0.1*	1.11*	0.11*	1.12*	0.10*	1.11*	0.10*	1.11*
	(0.01)	(0.01)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)
Education spending as % of GDP	-0.05*	-0.02	-0.33*	0.72*	-0.13	0.88	-0.15	0.86	-0.01	0.99

	(0.02)	(0.02)	(0.14)	(0.10)	(0.09)	(0.08)	(0.10)	(0.09)	(0.06)	(0.06)
Interaction terms										
Education spending as % of GDP * Primary	0.05*		0.31*	1.37*			0.20*	1.22*		
	(0.02)		(0.14)	(0.19)			(0.09)	(0.11)		
Middle East and North Africa * Education spending as % of GDP	0.05		0.31	1.36			0.26	1.30		
	(0.06)		(0.30)	(0.41)			(0.18)	(0.24)		
Year dummies (1986-2011)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Constant	2.00**	1.94**	8.55*	8.55*	7.85*	7.85*				
	(0.69)	(0.67)	(4.14)	(4.14)	(3.86)	(3.86)				
Cut 1 constant							-11.34***	-11.34***	-11.23***	-11.23***
							(3.03)	(3.03)	(2.91)	(2.91)
Cut 2 constant							-8.58**	-8.58**	-8.48**	-8.48**
							(3.04)	(3.04)	(2.92)	(2.92)
Cut 3 constant							-7.40*	-7.40*	-7.31*	-7.31*
							(3.02)	(3.02)	(2.91)	(2.91)
Cut 4 constant							-5.92	-5.92	-5.83*	-5.83*
							(3.03)	(3.03)	(2.91)	(2.91)
Cut 5 constant							-2.52	-2.52	-2.44	-2.44
							(3.03)	(3.03)	(2.92)	(2.92)
Observations	614	614	614	614	614	614	614	614	614	614
R squared	0.15	0.14								

Adjusted R squared	0.07	0.06								
Area under ROC curve			0.81	0.81	0.8	0.8				
Correctly specified			80%	80%	80.54%	80.54%				
BIC			-1133.9	-1133.9	-1124.3	-1124.3	2556.44	2556.44	2547.39	2547.39
Log-likelihood full model			-383.77	-383.77	-385.75	-385.75	-1081.47	-1081.47	-1083.73	-1083.73
McFadden's adjusted R squared			0.11	0.11	0.11	0.11	0.1	0.1	0.1	0.1
McKelvey & Zavoina’s R2			0.4	0.4	0.37	0.37	0.31	0.31	0.3	0.3

Note: OR_exponential coefficients(Odds ratio), Cluster Robust Standard errors in parentheses

Table 9 OLS regression results

	All education projects approved between 1996-2011						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable: Binary project ratings	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat
Project level characteristics							
Project length	0.07	0.08	0.09	0.01	0.01	0.01	0.00
	(0.07)	(0.08)	(0.07)	(0.01)	(0.01)	(0.01)	(0.01)
Project cost (logged, in USD millions)	0.01	0.00	0.02	0.02	0.01	0.01	0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Development Policy Loan	-0.06	-0.06	-0.00	0.04	0.08	0.05	0.01
	(0.10)	(0.11)	(0.11)	(0.10)	(0.10)	(0.09)	(0.07)
IDA project	0.02	0.02	-0.06	-0.06	-0.09	-0.07	-0.08
	(0.09)	(0.09)	(0.07)	(0.07)	(0.05)	(0.05)	(0.05)
Education sector	-0.10	-0.10	-0.05	-0.03	-0.01	0.01	0.01
	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)
Repeater project	0.06	0.05	0.06	0.06	0.06	0.06	0.08
	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Share of largest subsector in pct (logged)	-0.04	-0.03	-0.05	-0.04	-0.02	-0.04	-0.06
	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)
Emergency Recovery Loan	0.12	0.13	0.12	0.10	0.14	0.13	
	(0.12)	(0.12)	(0.10)	(0.11)	(0.11)	(0.10)	
Learning and Innovation Loan	-0.20	-0.21	-0.23	-0.21	-0.20	-0.21	
	(0.12)	(0.12)	(0.11)	(0.11)	(0.11)	(0.11)	

Specific Investment Loan	-0.06	-0.07	-0.09	-0.06	-0.02	-0.01
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Technical Assistance Loan	-0.31	-0.31	-0.34	-0.29	-0.30	-0.32
	(0.17)	(0.17)	(0.19)	(0.18)	(0.18)	(0.18)
Structural Adjustment Loan	-0.08	-0.08	-0.20	-0.17	-0.17	-0.16
	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)	(0.09)
Primary	-0.14	0.09	0.10	0.10	0.09	
	(0.11)	(0.06)	(0.05)	(0.06)	(0.05)	
Secondary	0.02	0.03	-0.00	-0.05	-0.04	
	(0.07)	(0.07)	(0.06)	(0.07)	(0.07)	
Tertiary	-0.03	-0.02	-0.02	0.00	0.01	
	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)	
Vocational	0.00	0.01	0.05	0.04	0.02	
	(0.11)	(0.11)	(0.10)	(0.10)	(0.10)	
General education	-0.07	-0.05	0.00	0.04	0.04	
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	
Regional dummies						
Africa	-0.02	-0.02	-0.07	-0.06		
	(0.11)	(0.11)	(0.10)	(0.09)		
East Asia Pacific	0.17	0.19*	0.10	0.12		
	(0.10)	(0.09)	(0.10)	(0.09)		
South Asia	0.13	0.15	0.13	0.11		
	(0.10)	(0.10)	(0.10)	(0.09)		
Latin America and Caribbean	0.12	0.12	0.01	0.02		
	(0.12)	(0.13)	(0.11)	(0.10)		
Middle East and North Africa	-0.07	0.10	-0.08	-0.09		
	(0.45)	(0.15)	(0.12)	(0.11)		
Country level characteristics						

Primary school gross enrollment (logged)	-0.31*	-0.32*					
	(0.14)	(0.13)					
GDP growth (annual %)	0.01	0.02					
	(0.01)	(0.01)					
Education spending as % of GDP	-0.08*	-0.04*					
	(0.03)	(0.02)					
Government effectiveness	0.15*	0.14*					
	(0.07)	(0.06)					
Interaction terms							
Education spending*Primary	0.06*						
	(0.03)						
East Asia Pacific*Government effectiveness	0.01						
	(0.09)						
Middle East and North Africa*Education spending	0.03						
	(0.07)						
Year dummies	Y	Y	Y	Y	Y	Y	Y
Constant	2.60***	2.50***	0.94	0.76	0.75	0.87	0.96
	(0.71)	(0.68)	(0.29)	(0.28)	(0.26)	(0.22)	(0.21)
Observations	431	431	526	526	526	526	526
R squared	0.19	0.18	0.13	0.07	0.05	0.04	0.02
Adjusted R squared	0.1	0.09	0.06	0.03	0.02	0.02	0.001

Cluster robust standard errors in parentheses *p<0.05 ** p<0.01 ***p<0.001

Table 10 Logistic regression results (coefficients and odds ratio)

	All education projects approved between 1996-2011					
	(1)	(1)_OR	(2)	(2)_OR	(3)	(3)_OR
Dependent variable: Binary project ratings	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat
Project level characteristics						
Project length	0.38	1.46	0.40	1.50	0.50	1.65
	(0.41)	(0.59)	(0.41)	(0.62)	(0.41)	(0.68)
Project cost (logged, in USD millions)	0.04	1.04	0.03	1.03	0.10	1.11
	(0.13)	(0.14)	(0.12)	(0.13)	(0.09)	(0.10)
Development Policy Loan	-0.39	0.68	-0.50	0.61	-0.07	0.93
	(0.73)	(0.49)	(0.77)	(0.47)	(0.64)	(0.59)
IDA project	0.14	1.15	0.11	1.12	-0.39	0.68
	(0.59)	(0.68)	(0.58)	(0.65)	(0.45)	(0.30)
Education sector	-0.60	0.55	-0.58	0.56	-0.21	0.81
	(0.36)	(0.20)	(0.36)	(0.20)	(0.35)	(0.29)
Repeater project	0.41	1.50	0.38	1.47	0.40	1.49
	(0.27)	(0.41)	(0.27)	(0.40)	(0.22)	(0.33)
Share of largest subsector in pct (logged)	-0.30	0.74	-0.21	0.81	-0.36	0.69
	(0.39)	(0.29)	(0.38)	(0.31)	(0.34)	(0.23)
Emergency Recovery Loan	1.00	2.71	1.02	2.77	0.80	2.21
	(0.82)	(2.21)	(0.83)	(2.31)	(0.65)	(1.45)
Learning and Innovation Loan	-1.06	0.35	-1.11	0.33	-1.14	0.32
	(0.69)	(0.24)	(0.69)	(0.23)	(0.58)	(0.18)
Specific Investment Loan	-0.42	0.66	-0.47	0.62	-0.53	0.59
	(0.48)	(0.32)	(0.49)	(0.30)	(0.42)	(0.25)
Technical Assistance Loan	-1.61*	0.20*	-1.58	0.21	-1.68	0.19
	(0.80)	(0.16)	(0.83)	(0.17)	(0.89)	(0.17)

Structural Adjustment Loan	-0.41	0.66	-0.35	0.70	-1.07	0.34
	(0.45)	(0.30)	(0.45)	(0.32)	(0.45)	(0.15)
Primary	-1.02	0.36	0.58	1.79	0.52	1.69
	(0.79)	(0.29)	(0.34)	(0.61)	(0.31)	(0.52)
Secondary	0.12	1.12	0.16	1.17	-0.07	0.93
	(0.41)	(0.46)	(0.41)	(0.48)	(0.39)	(0.36)
Tertiary	-0.16	0.85	-0.13	0.88	-0.13	0.88
	(0.47)	(0.40)	(0.47)	(0.41)	(0.43)	(0.38)
Vocational	0.02	1.02	0.03	1.03	0.29	1.33
	(0.68)	(0.70)	(0.67)	(0.69)	(0.55)	(0.73)
General education	-0.38	0.68	-0.27	0.76	0.01	1.01
	(0.36)	(0.24)	(0.35)	(0.27)	(0.35)	(0.35)
Regional dummies						
Africa	-0.04	0.96	-0.02	0.98	-0.28	0.75
	(0.56)	(0.54)	(0.56)	(0.55)	(0.52)	(0.39)
East Asia Pacific	1.85**	6.35**	1.55**	4.70**	0.80	2.22
	(0.67)	(4.29)	(0.53)	(2.50)	(0.62)	(1.38)
South Asia	0.89	2.44	1.00	2.71	0.89	2.44
	(0.60)	(1.46)	(0.61)	(1.66)	(0.59)	(1.44)
Latin America and Caribbean	0.75	2.12	0.70	2.02	0.11	1.11
	(0.65)	(1.38)	(0.65)	(1.31)	(0.63)	(0.70)
Middle East and North Africa	-0.47	0.63	0.66	1.93	-0.38	0.68
	(2.41)	(1.51)	(0.82)	(1.58)	(0.63)	(0.43)
Country level characteristics						
Primary school gross enrollment (logged)	-1.68*	0.19*	-1.73*	0.18*		
	(0.75)	(0.14)	(0.70)	(0.12)		
GDP growth (annual %)	0.11*	1.12*	0.11*	1.12*		
	(0.05)	(0.06)	(0.05)	(0.06)		
Education spending as % of GDP	-0.48*	0.62*	-0.25*	0.77*		
	(0.19)	(0.12)	(0.12)	(0.10)		
Government effectiveness	0.96*	2.61*	0.90*	2.46*		

	(0.48)	(1.25)	(0.44)	(1.08)		
Interaction terms						
Education spending*Primary	0.37*	1.45*				
	(0.18)	(0.26)				
East Asia Pacific*Government effectiveness	1.09	2.97				
	(0.88)	(2.62)				
Middle East and North Africa*Education spending	0.23	1.25				
	(0.37)	(0.47)				
Year dummies	Y	Y	Y	Y	Y	Y
Constant	11.78**		10.92**		2.61	
	(4.11)		(3.82)		(1.87)	
Observations	431	431	431	431	526	526
Area Under ROC curve	0.83		0.83		0.81	
Correctly Specified	82.79%		82.50%		78.07%	
BIC	864.411		848.372		1164.844	
Log-likelihood full model	-282.401		-284.152		-448.194	
McFadden’s adjusted R squared	0.13		0.13		0.14	
McKelvey & Zavoina's R squared	0.41		0.4		0.34	

Cluster robust standard errors in parentheses *p<0.05 ** p<0.01 ***p<0.001
OR: Exponentiated coefficients

Table 11 Regression results for ordered project outcome

	All education projects approved between 1996-2011							
	(1)	(2)	(3)	(3)_OR	(4)	(4)_OR	(5)	(5)_OR
	OLS	OLS	Ologit	Ologit_OR	Ologit	Ologit_OR	Ologit	Ologit_OR
Dependent variable: Six ordinal project outcomes	0-5 rating	0-5 rating	0-5 rating	0-5 rating	0-5 rating	0-5 rating	0-5 rating	0-5 rating
Project level characteristics								
Project length	0.18	0.20	0.58	1.79	0.61	1.84	0.62	1.86
	(0.21)	(0.22)	(0.38)	(0.68)	(0.39)	(0.72)	(0.34)	(0.63)
Project cost (logged, in USD millions)	-0.04	-0.04	-0.09	0.92	-0.10	0.91	-0.02	0.98
	(0.04)	(0.04)	(0.09)	(0.09)	(0.09)	(0.08)	(0.09)	(0.08)
Development Policy Loan	0.10	0.06	-0.05	0.95	-0.18	0.83	0.25	1.28
	(0.25)	(0.27)	(0.51)	(0.49)	(0.53)	(0.44)	(0.49)	(0.62)
IDA project	-0.04	-0.06	-0.11	0.90	-0.18	0.83	-0.53	0.59
	(0.17)	(0.18)	(0.33)	(0.30)	(0.36)	(0.30)	(0.27)	(0.16)
Education sector	-0.19	-0.19	-0.34	0.71	-0.34	0.71	-0.14	0.87
	(0.14)	(0.14)	(0.29)	(0.20)	(0.30)	(0.21)	(0.31)	(0.27)
Repeater project	0.00	-0.02	-0.14	0.87	-0.17	0.85	-0.07	0.93
	(0.10)	(0.10)	(0.20)	(0.17)	(0.20)	(0.17)	(0.19)	(0.17)
Share of largest subsector in pct (logged)	-0.13	-0.08	-0.17	0.84	-0.10	0.90	-0.21	0.81
	(0.12)	(0.12)	(0.24)	(0.21)	(0.25)	(0.22)	(0.24)	(0.20)
Emergency Recovery Loan	0.41	0.44	1.03	2.81	1.05	2.85	0.84	2.32
	(0.28)	(0.29)	(0.60)	(1.68)	(0.62)	(1.77)	(0.55)	(1.28)
Learning and Innovation Loan	-0.27	-0.32	-0.54	0.58	-0.66	0.52	-0.73	0.48

	(0.26)	(0.26)	(0.54)	(0.32)	(0.51)	(0.27)	(0.52)	(0.25)
Specific Investment Loan	0.00	-0.02	0.07	1.07	0.04	1.04	-0.18	0.84
	(0.16)	(0.16)	(0.28)	(0.30)	(0.28)	(0.29)	(0.27)	(0.23)
Technical Assistance Loan	-0.53	-0.52	-0.90	0.41	-0.87	0.42	-1.09	0.34
	(0.31)	(0.33)	(0.58)	(0.24)	(0.63)	(0.26)	(0.70)	(0.24)
Structural Adjustment Loan	-0.64**	-0.60**	-1.07*	0.34*	-0.97*	0.38*	-1.22	0.29
	(0.23)	(0.23)	(0.42)	(0.14)	(0.42)	(0.16)	(0.37)	(0.11)
Primary	-0.47	0.20	-0.84	0.43	0.41	1.51	0.44	1.56
	(0.26)	(0.13)	(0.55)	(0.24)	(0.27)	(0.40)	(0.26)	(0.40)
Secondary	0.19	0.21	0.48	1.62	0.49	1.63	0.28	1.33
	(0.16)	(0.15)	(0.30)	(0.49)	(0.30)	(0.49)	(0.29)	(0.39)
Tertiary	0.14	0.18	0.42	1.53	0.50	1.64	0.44	1.55
	(0.20)	(0.19)	(0.41)	(0.63)	(0.40)	(0.65)	(0.37)	(0.58)
Vocational	0.21	0.22	0.38	1.47	0.41	1.50	0.49	1.62
	(0.24)	(0.24)	(0.50)	(0.73)	(0.48)	(0.72)	(0.39)	(0.64)
General education	0.08	0.14	0.33	1.40	0.43	1.53	0.45	1.56
	(0.16)	(0.16)	(0.33)	(0.47)	(0.34)	(0.52)	(0.31)	(0.49)
Regional dummies								
Africa	-0.02	-0.02	0.05	1.05	0.07	1.07	-0.07	0.93
	(0.20)	(0.21)	(0.39)	(0.41)	(0.40)	(0.43)	(0.34)	(0.32)
East Asia Pacific	0.56**	0.47*	1.44**	4.22**	1.10*	3.00*	0.73	2.07
	(0.21)	(0.22)	(0.49)	(2.08)	(0.48)	(1.45)	(0.47)	(0.98)
South Asia	0.24	0.30	0.50	1.65	0.61	1.84	0.60	1.82
	(0.17)	(0.19)	(0.33)	(0.55)	(0.36)	(0.67)	(0.37)	(0.68)
Latin America and Caribbean	0.27	0.26	0.63	1.88	0.57	1.77	0.08	1.09
	(0.24)	(0.25)	(0.49)	(0.91)	(0.51)	(0.90)	(0.41)	(0.44)
Middle East and North Africa	-0.59	0.14	-1.22	0.29	0.29	1.34	-0.34	0.71
	(0.63)	(0.28)	(1.21)	(0.36)	(0.55)	(0.74)	(0.41)	(0.29)

Country level characteristics								
Primary school gross enrollment (logged)	-0.59	-0.71*	-0.98	0.37	-1.22*	0.30*		
	(0.31)	(0.29)	(0.59)	(0.22)	(0.54)	(0.16)		
GDP growth (annual %)	0.03	0.04	0.07	1.08	0.08	1.08		
	(0.02)	(0.02)	(0.04)	(0.05)	(0.04)	(0.05)		
Education spending as % of GDP	-0.17**	-0.06	-0.34**	0.71**	-0.13	0.88		
	(0.06)	(0.04)	(0.12)	(0.08)	(0.08)	(0.07)		
Government effectiveness	0.23	0.23	0.47	1.60	0.48	1.62		
	(0.15)	(0.14)	(0.30)	(0.47)	(0.29)	(0.46)		
Interaction terms								
Education spending*Primary	0.16**		0.30*	1.34*				
	(0.06)		(0.12)	(0.16)				
East Asia Pacific*Government effectiveness	0.55*		1.45*	4.25*				
	(0.27)		(0.61)	(2.61)				
Middle East and North Africa*Education spending	0.14		0.29	1.34				
	(0.10)		(0.21)	(0.28)				
Year dummies	Y	Y	Y	Y	Y	Y	Y	Y
Constant	7.12***	7.18***						
	(1.65)	(1.56)						
cut1								
Constant			-12.26***		-12.43***		-6.19	
			(3.26)		(3.03)		(1.32)	
cut2								
Constant			-10.00**		-10.19***		-4.13	

					(3.19)		(2.96)		(1.31)	
cut3										
Constant					-8.56**		-8.76**		-2.74	
					(3.18)		(2.95)		(1.29)	
cut4										
Constant					-6.95*		-7.16*		-1.12	
					(3.19)		(2.96)		(1.27)	
cut5										
Constant					-3.15		-3.46		2.56	
					(3.20)		(2.96)		(1.26)	
Observations	431		431		431	431	431	431	526	526
R squared		0.19		0.18						
Adjusted R squared		0.11		0.09						
BIC					1902.125	1902.125	1893.735	1893.735	2615.118	2615.118
Log-likelihood full model					-788.232	-788.232	-793.806	-793.806	-1159.565	-1159.565
McFadden's adjusted R squared					0.08	0.08	0.07	0.07	0.08	0.08
McKelvey & Zavoina’s R squared					0.35	0.35	0.34	0.34	0.3	0.3

Cluster robust standard errors in parentheses *p<0.05 ** p<0.01 ***p<0.001
OR: Exponentiated coefficients

Table 12 OLS and logistic regression results with country fixed effects added

	All education projects approved between 1996-2011							
	(1)	(2)	(3)	(3)_OR	(4)	(4)_OR	(5)	(5)_OR
	OLS	OLS	Logit	Logit_OR	Logit	Logit_OR	Logit	Logit_OR
Dependent variable: Binary project ratings	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat
Project level characteristics								
Project length	0.18*	0.19*	0.96	2.60	1.06*	2.88*	1.14	3.13
	(0.08)	(0.08)	(0.52)	(1.36)	(0.53)	(1.53)	(0.52)	(1.63)
Project cost (logged, in USD millions)	0.01	0.01	0.11	1.12	0.08	1.08	0.20	1.22
	(0.02)	(0.02)	(0.18)	(0.20)	(0.17)	(0.19)	(0.13)	(0.16)
Development Policy Loan	-0.06	-0.07	-0.28	0.75	-0.39	0.68	-0.45	0.64
	(0.12)	(0.12)	(1.03)	(0.78)	(1.05)	(0.71)	(0.85)	(0.54)
IDA project	-0.02	-0.07	-0.12	0.89	-0.55	0.57	-2.66	0.07
	(0.16)	(0.16)	(2.77)	(2.46)	(3.17)	(1.82)	(1.66)	(0.12)
Education sector	-0.08	-0.08	-0.57	0.56	-0.60	0.55	-0.16	0.86
	(0.07)	(0.07)	(0.55)	(0.31)	(0.54)	(0.30)	(0.47)	(0.40)
Repeater project	0.05	0.05	0.52	1.67	0.51	1.66	0.40	1.49
	(0.05)	(0.05)	(0.43)	(0.72)	(0.43)	(0.71)	(0.35)	(0.52)
Share of largest subsector in pct (logged)	-0.07	-0.06	-0.54	0.58	-0.35	0.70	-0.58	0.56
	(0.06)	(0.06)	(0.57)	(0.33)	(0.54)	(0.38)	(0.44)	(0.25)
Emergency Recovery Loan	0.11	0.12	3.27	26.44	2.13	8.43	0.93	2.52
	(0.16)	(0.16)	(2.28)	(60.34)	(1.62)	(13.65)	(1.06)	(2.68)
Learning and Innovation Loan	-0.26*	-0.25*	-1.55	0.21	-1.32	0.27	-1.68	0.19

	(0.12)	(0.12)	(1.00)	(0.21)	(0.97)	(0.26)	(0.86)	(0.16)
Specific Investment Loan	-0.13	-0.12	-1.10	0.33	-1.02	0.36	-0.89	0.41
	(0.08)	(0.08)	(0.66)	(0.22)	(0.67)	(0.24)	(0.61)	(0.25)
Technical Assistance Loan	-0.50*	-0.49*	-16.71	0.00	-16.38	0.00	-15.71	0.00
	(0.20)	(0.20)	(1219.22)	(0.00)	(947.11)	(0.00)	(894.24)	(0.00)
Structural Adjustment Loan	-0.09	-0.09	-0.70	0.50	-0.61	0.54	-0.92	0.40
	(0.12)	(0.12)	(1.04)	(0.52)	(1.01)	(0.55)	(0.71)	(0.28)
Primary	-0.16	0.08	-1.57	0.21	0.85	2.34	0.69	2.00
	(0.15)	(0.07)	(1.31)	(0.27)	(0.52)	(1.22)	(0.46)	(0.92)
Secondary	0.03	0.03	0.20	1.23	0.20	1.22	-0.05	0.95
	(0.08)	(0.08)	(0.62)	(0.76)	(0.61)	(0.75)	(0.53)	(0.50)
Tertiary	-0.07	-0.07	-0.40	0.67	-0.39	0.68	-0.64	0.53
	(0.08)	(0.08)	(0.60)	(0.40)	(0.60)	(0.41)	(0.51)	(0.27)
Vocational	0.05	0.05	0.20	1.23	0.12	1.13	0.20	1.22
	(0.10)	(0.10)	(0.97)	(1.19)	(0.91)	(1.03)	(0.74)	(0.91)
General education	-0.03	-0.02	-0.08	0.92	-0.01	0.99	0.17	1.19
	(0.07)	(0.07)	(0.51)	(0.47)	(0.51)	(0.50)	(0.43)	(0.52)
Country level characteristics								
Primary school gross enrollment (logged)	-0.15	-0.13	-0.98	0.37	-0.58	0.56		
	(0.32)	(0.32)	(2.61)	(0.98)	(2.56)	(1.44)		
GDP growth (annual %)	0.01	0.01	0.12	1.13	0.11	1.11		
	(0.01)	(0.01)	(0.12)	(0.13)	(0.12)	(0.13)		
Education spending as % of GDP	-0.07	-0.05	-0.75*	0.47*	-0.48	0.62		
	(0.04)	(0.04)	(0.38)	(0.18)	(0.35)	(0.21)		
Government effectiveness	0.11	0.09	1.10	2.99	1.16	3.18		
	(0.24)	(0.24)	(2.02)	(6.05)	(2.02)	(6.42)		

Interaction terms								
Education spending*Primary	0.06		0.57*	1.77*				
	(0.03)		(0.29)	(0.51)				
Selected Year dummies								
Year 2008	-0.63	-0.66	-3.02***	0.05***	-3.11***	0.04***	-2.91	0.05
	(0.35)	(0.36)	(0.91)	(0.04)	(0.89)	(0.04)	(0.81)	(0.04)
Year 2009	-0.28	-0.30	0.05	1.06	-0.21	0.81	-0.70	0.50
	(0.30)	(0.30)	(0.99)	(1.04)	(0.98)	(0.79)	(0.87)	(0.43)
Year 2010	0.20	0.13	-0.15	0.86	-0.10	0.90	-0.29	0.75
	(0.32)	(0.33)	(1.00)	(0.86)	(1.00)	(0.91)	(0.87)	(0.66)
Year 2011	-0.48	-0.49	-2.66*	0.07*	-3.00*	0.05*	-3.00	0.05
	(0.38)	(0.39)	(1.29)	(0.09)	(1.27)	(0.06)	(1.12)	(0.06)
Year dummies	Y	Y	Y	Y	Y	Y	Y	Y
Constant	2.23	2.04						
	(1.46)	(1.47)						
Observations	431	431	312	312	312	312	391	391
Number of countries	52	52	52	52	52	52	61	61

Standard errors in parentheses *p<0.05 ** p<0.01 ***p<0.001

Coefficients and standard errors for year 1996-2011 are not shown in the table

Table 13 Logistic regression results by subsectors

	All projects from 1986-2011				
	(1)	(2)	(3)	(4)	(5)
	Primary	Secondary	Tertiary	Vocational	General Education
Depedent variable: Binary project ratings	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat
Project level characteristics					
Project length	0.03	0.21	0.17	-0.41	-0.02
	(0.11)	(0.14)	(0.19)	(0.34)	(0.23)
Project cost (logged, in USD millions)	0.30*	0.41	-0.42	-0.46	-0.37
	(0.15)	(0.28)	(0.49)	(0.47)	(0.27)
Development Policy Loan	1.23	0.00	0.00	0.00	-0.46
	(1.12)	(.)	(.)	(.)	(1.09)
IDA project	0.45	-0.39	-0.00	-4.38	-0.47
	(0.67)	(0.90)	(0.85)	(2.17)	(0.84)
Education sector	-1.10	0.00	0.08	2.49	-0.56
	(0.67)	(.)	(0.63)	(1.32)	(0.85)
Repeater project	-0.26	-0.02	-0.50	-0.93	-0.50
	(0.39)	(0.64)	(0.63)	(1.78)	(0.58)
Share of largest subsector in pct (logged)	0.47	-1.08	-0.07	-2.29	-0.37
	(0.32)	(0.85)	(0.58)	(1.33)	(0.63)
Regional dummies					
Africa	-0.90	-1.75	-0.10	-0.32	0.24
	(0.74)	(1.02)	(1.05)	(1.88)	(0.95)
East Asia Pacific	1.00	1.52	3.36**	1.87	0.32

	(0.78)	(0.87)	(1.03)	(2.20)	(1.40)
South Asia	0.75	0.00	0.88	1.89	1.03
	(0.71)	(.)	(1.31)	(2.95)	(0.91)
Latin America and Caribbean	0.38	0.53	2.51*	-2.40	0.34
	(0.81)	(1.00)	(1.12)	(1.48)	(0.86)
Middle East and North Africa	-0.61	10.19*	-0.15	-6.26	-64.61***
	(3.27)	(4.68)	(3.31)	(4.13)	(3.61)
Country level characteristics					
Primary school gross enrollment (logged)	0.17	-2.25	0.41	-3.36	-1.71
	(0.71)	(2.22)	(1.46)	(2.31)	(1.61)
GDP growth (annual %)	0.02	0.07	0.03	0.00	0.22
	(0.08)	(0.15)	(0.14)	(0.31)	(0.12)
Education spending as % of GDP	-0.01	0.22	0.07	-0.21	-0.46*
	(0.08)	(0.12)	(0.19)	(0.39)	(0.23)
Interaction terms					
Middle East and North Africa*Education spending	0.13	-1.43*	0.29	1.00	8.67***
	(0.46)	(0.69)	(0.48)	(0.72)	(0.50)
Year dummies	Y	Y	Y	Y	Y
Constant	-2.30	11.08	-0.71	33.31	13.59
	(3.73)	(9.65)	(8.32)	(15.85)	(7.37)
Observations	217	96	121	60	108
Area Under ROC curve	0.75	0.77	0.86	0.92	0.84
Correctly specified	76.35%	84.55%	84.62%	81.70%	85.80%
McKelvey & Zavoina	0.3	0.44	0.43	0.74	0.87

Cluster robust standard errors in parenthesis * p<0.05 ** p<0.01 *** p<0.001

Table 14 Logistic regression results with additional project and governance measures added

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable: Binary project ratings	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat
Government effectiveness	2.15**	2.27*	1.09*	1.45*	1.04	2.17**	1.14	1.01
	(0.75)	(1.01)	(0.46)	(0.69)	(0.60)	(0.67)	(0.92)	(0.81)
Counterpart funding (% of government funding out of total commitment)	0.01							
	(0.02)							
Staff supervision weeks (% of supervision out of total weeks)		0.02						
		(0.02)						
EFA project			0.12					
			(0.33)					
Perceptions on voice and accountability				-0.41				
				(0.37)				
Perception on political stability and violence					0.03			
					(0.26)			
Regulatory Quality						-1.25*		
						(0.58)		
Rule of Law							-0.06	
							(0.60)	

Control of corruption								0.07
								(0.55)
Constant	11.93	17.53	10.67**	13.23**	13.34**	12.98**	13.33**	13.21**
	(8.66)	(11.71)	(3.96)	(4.27)	(4.34)	(4.28)	(4.36)	(4.64)
Observations	185	139	481	418	418	418	418	418
Area Under ROC curve	0.95	0.97	0.85	0.87	0.87	0.88	0.87	0.87
Correctly Specified	91.14%	90.88%	81.60%	85.99%	86.13%	86.41%	86.41%	86.13%
McFadden's adjusted R squared	0.39	0.41	0.24	0.25	0.24	0.25	0.24	0.24
McKelvey & Zavoina's R squared	0.8	0.86	0.52	0.55	0.54	0.55	0.54	0.54

Cluster robust standard errors in parentheses *p<0.05 ** p<0.01 ***p<0.001

Table 15 Logistic regression results with additional country characteristics added

	(1)	(2)	(3)	(4)
Dependent variable: binary project ratings	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat
Government effectiveness	1.01	0.74	0.96*	1.05*
	(0.53)	(0.59)	(0.47)	(0.46)
Measure of democracy/Polity2 (-10 monarchy +10 democracy)	-0.05			
	(0.04)			
State Fragility Index (0 lowest fragility 23 highest fragility)		-0.04		
		(0.06)		
Population (logged)			-0.01	
			(0.12)	
Net ODA (logged, constant USD 2011)				-0.28
				(0.17)
Constant	10.44*	11.02**	10.79*	17.04***
	(4.20)	(4.15)	(4.44)	(4.95)
Observations	441	445	445	476
Area Under ROC curve	0.86	0.85	0.85	0.88
Correctly Specified	83.24%	82.80%	82.66%	84.43%
McFadden's adjusted R squared	0.25	0.25	0.25	0.29
McKelvey & Zavoina's R squared	0.53	0.51	0.52	0.6

Cluster robust standard errors in parentheses *p<0.05 ** p<0.01 ***p<0.001

Note: Project length, cost, dummies for DPL, IDA, repeater, education sector, regional dummies, type of loans, subsector, primary school gross enrollment, GDP growth, education spending as % of GDP, interaction variables and year dummies were added in the regression but omitted in the table

Table 16 Marginal effects from ordinal logistic regression

Probability for each category of outcome	Marginal effects	Primary education	Marginal effects	Secondary education	Marginal effects	Tertiary education	Marginal effects
The probability to receive "highly unsatisfactory" rating	0.006	The probability to receive "highly unsatisfactory rating" for a primary education project	0.01	The probability to receive "highly unsatisfactory rating" for a secondary education project	0.004	The probability to receive "highly unsatisfactory rating" for a tertiary education project	0.004
	(0.003)		(0.01)		(0.002)		(0.003)
The probability to receive "unsatisfactory" rating	0.057	The probability to receive "unsatisfactory" rating for a primary education project	0.098	The probability to receive "unsatisfactory" rating for a secondary education project	0.04	The probability to receive "unsatisfactory" rating for a tertiary education project	0.04
	(0.01)		(0.036)		(0.01)		(0.02)
The probability to receive "moderately unsatisfactory" rating	0.16	The probability to receive "moderately unsatisfactory" rating for a primary education project	0.24	The probability to receive "moderately unsatisfactory" rating for a secondary education project	0.13	The probability to receive "moderately unsatisfactory" rating for a tertiary education project	0.13
	(0.02)		(0.06)		(0.03)		(0.04)
The probability to receive "moderately satisfactory" rating	0.37	The probability to receive "moderately satisfactory" rating for a primary education project	0.38	The probability to receive "moderately satisfactory" rating for a secondary education project	0.34	The probability to receive "moderately satisfactory" rating for a tertiary education project	0.34
	(0.03)		(0.03)		(0.04)		(0.05)
The probability to receive "satisfactory" rating	0.39	The probability to receive "satisfactory" rating for a primary education project	0.27	The probability to receive "satisfactory" rating for a secondary education project	0.46	The probability to receive "satisfactory" rating for a tertiary education project	0.46

	(0.03)		(0.08)		(0.07)	(0.09)
The probability to receive "highly satisfactory" rating	0.02	The probability to receive "highly satisfactory" rating for a primary education project	0.01	The probability to receive "highly satisfactory" rating for a secondary education project	0.02	The probability to receive "highly satisfactory" rating for a tertiary education project	0.02
	(0.01)		(0.01)		(0.01)	(0.01)

Note: Standard errors in parenthesis

Note: All other variables set at their mean values

Research Question 2: Perceived factors for education project outcomes

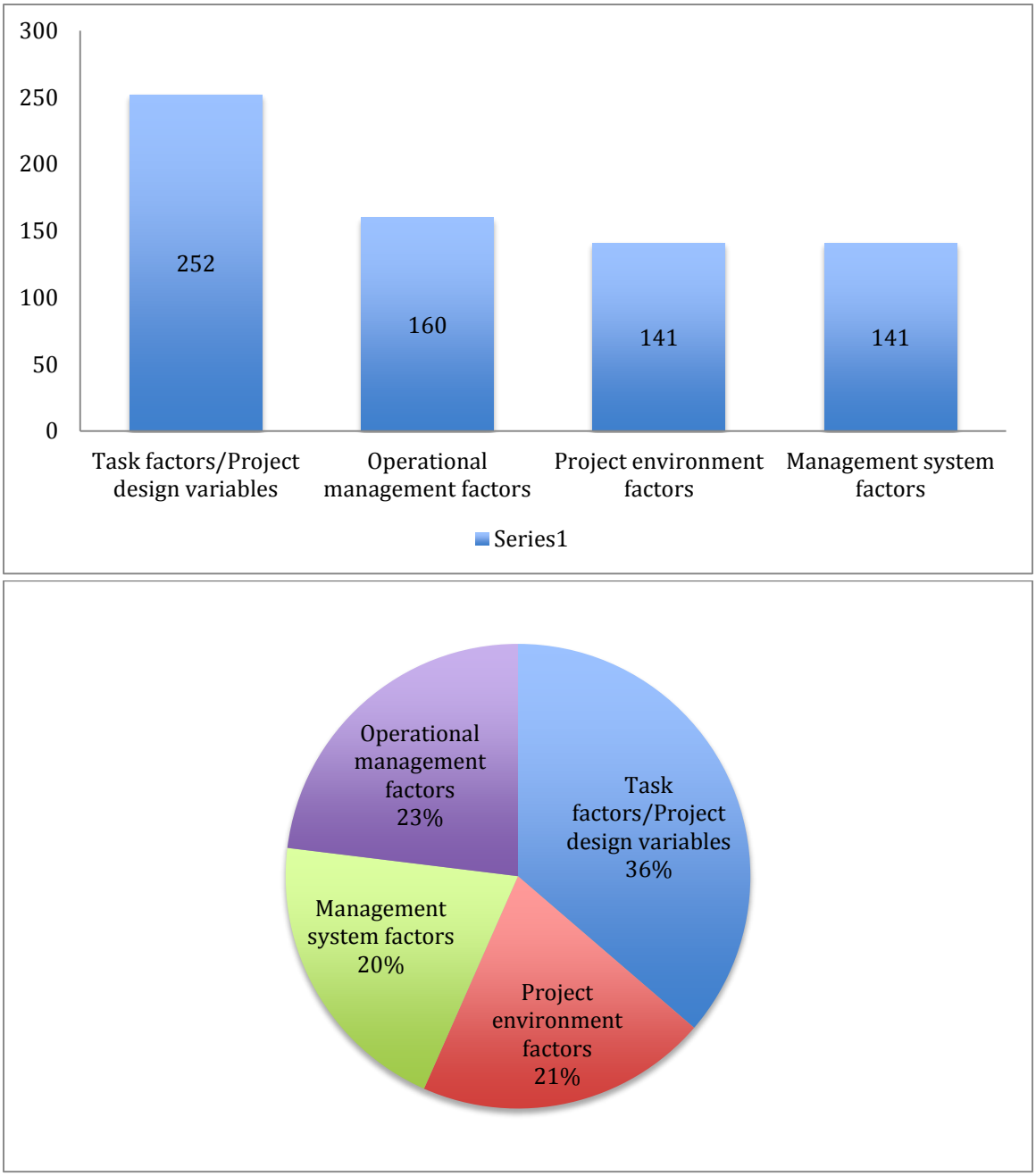


Figure 21 Frequency of words in categories

Figure 18 illustrates the frequency of factors mentioned by the interviewees, categorized by the analytical concept terms used by Middleton (1985). Project design factors are the most frequently mentioned (36%) by the interviewees. These factors include project goals, design, institution building needs, and political consensus about the project being implemented in the country. Operational management factors are the next most frequently mentioned (23%) by the interviewees. Operational management factors include factors that are associated with input management (such as disbursement, procurement and technical assistance), monitoring, and the nature of the relationship between the borrower and the Bank during implementation of the project.

Project environment factors and management factors are followed by operational management factors (20–21%) by the interviewees. Project environment factors include policy stability (or instability), corruption, and bureaucratic procedures. Management system factors include factors related to issues of the structure and functions of project implementing agency. Examples of management system factors include organizational issues (such as different priorities for mid-level and high-level officers, conflicts between different ministries), staffing and skills of the Bank and the local staff, government implementation capacity and the continuity of the key staff.

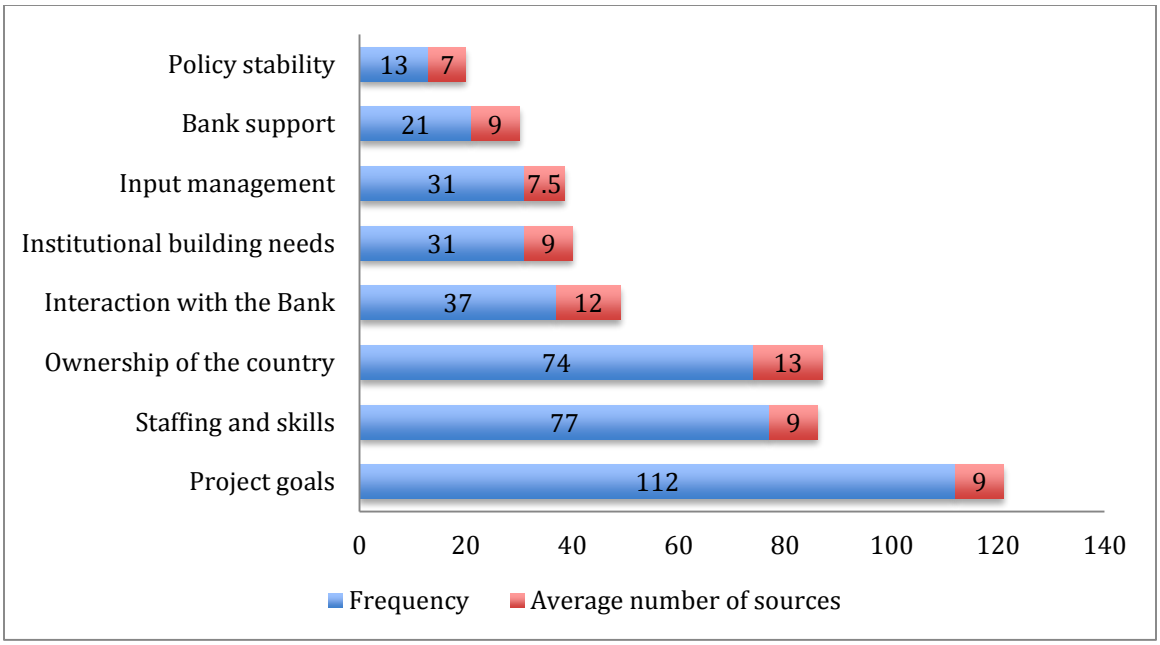


Figure 22 Frequency and average number of sources by sub-categories

Figure 19 shows the frequency and the average number of interviewees (sources) of each sub-category that are most frequently mentioned by the interviewees.

Theme 1: Project goals/design factor

Under the sub-category of "project goals," are smaller categories of "focused target," "good design," "Design issues (consists of complicated project development objectives, not aligning with government priorities and objectives)," "being too ambitious," and "various partners involved."

Focused target – not being too ambitious in project objectives

"Focused target" mainly had two meanings. Focused activities and project objectives, and focused target population who were the beneficiaries of the project.

Focused project with less activities were much more likely to show better performance than projects with too many activities as the government capability is limited (Interview #7[30]). Also, instead of implementing the project at the nation-wide level, interviewees noted that projects tended to show better results when they limit the target to a certain population or area.

> It worked well because it was focused, it was done just in rural schools. They didn't try to do all schools. It worked really well, because it was focused. (Interview #5[31]).

Reducing the number of schools that participate in the project, and focusing on schools in the rural areas instead of schools nation-wide, was considered as part of a reason for successful implementation of projects.

Simplified project goals also mean not to be too ambitious with project objectives at the preparation stage of the project. Interviewees also mentioned that projects that had goals that were too ambitious (mentioned 19 times) were less likely to receive good evaluation ratings. An ambitious project can be described in terms of project design and in terms of feasibility. An ambitious project would have too many objectives and have limited time frame to implement all the objectives. For example, a project with objectives of bringing primary education to all age group, and providing schoolings to all girls within a certain time would be too ambitious if the country's population is growing very fast (Interview #6[32]). Also, an ambitious project would have objectives that are not realistic, therefore, having difficulties in logistically carrying out the project. If the

[30] Interview #7 (May 5, 2014)
[31] Interview #5 (March 12, 2014)
[32] Interview #6 (May 27, 2014)

project aims to build schools while there are no roads to carry the materials to build schools, it will not be a feasible project objective (Interview #6[33]).

Focused target with less ambitious project design may be the factors for better evaluation ratings. Since ratings are based on to what extent the project has reached its objectives, the less number of goals and the lower the target indicators, the more likely the project will reach its objectives, thus receive higher ratings. Thus, the more ambitious project objectives are, the more likely the project will receive lower ratings.

Project Design

Project design was another sub-category under the project goals/design factors. Although "focused target" factor, that is discussed above is part of project design, the "project design" factor here has a broader definition. A project that is well designed would be appropriate in the local context, taking both the micro and macro situations, and include components that truly mattered to the recipients (Interview #6, #13[34]). A well-designed project would also be designed by a person who has expertise in education, as he/she would be well aware of the learning processes in classrooms and be able to pay attention to the details of the project (Interview #5[35]). The IEG measures the quality of project design, which is represented by the term "Quality at entry." The quality at entry measures whether the design of the project was appropriate for the setting in order to achieve the outcomes (Interview #8[36]).

---

33 Interview #6 (May 27, 2014)
34 Interview #6 (May 27, 2014), Interview #13 (May 7, 2014)
35 Interview #5 (March 12, 2014)
36 Interview #8 (May 13, 2014)

On the other hand, some of the "design issues" mentioned by the interviewees had complicated project development objectives (goals), being too innovative, and had project objectives that do not align with recipient country's government objectives and priorities. Complicated project development objectives, in part, overlapped with "too ambitious" projects and projects with less focused targets. Other than the overlapped definition, complicated objectives would also have difficulties in balancing equity and efficiency (Interview #3[37]). Complicated project objectives would also have many different components not directly related to each other, or that in order for one component to succeed the other three components need to work well (Interview #8[38]).

However, the "project design" factor needs careful interpretation. The interviewees underlined that it is important to note that a well-design project itself will not lead to better project performance. Other factors at the management and operational side can play a decisive role in project performance. A project with good quality design would not have high performance if any of the factors at the implementation part are not satisfactorily met.

> So once the concept is accepted and the Bank's resources are committed to preparing the project then it's as if the horses leaves the gates (Interview #9[39])
>
> I think most of it is the country side, assuming that we designed it correctly, the rest is up to the country (Interview #11[40])

I discuss the factors at the implementation level later in this section.

---

[37] Interview #3 (April 21, 2014)
[38] Interview #8 (May 13, 2014)
[39] Interview #9 (May 13, 2014)
[40] Interview #11 (March 11, 2014)

Various partners involved

The interviewees also noted that if there are more partners involved in the same project, then the objectives tend to become complicated because the interests of all the donor agencies needs to be reflected in the objectives. One of the interviewees particularly pointed out that when multiple donor agencies are involved, the project implementation risk increases (Interview #7[41]). The project will confront difficulties when one of the donors decides to stop working together (Interview #9[42]). Maintaining policy dialogues among different donor agencies is also a difficult task (Interview #15[43]). Thus, the factor of "various donors involved" is associated with both having complicated objectives and also the project outcomes.

Theme 2: Government capacity, staffing and skills (Management system factor)

Government/institutional capacity

The second theme that emerged from the data is government capacity, staffing and skills. "Government capacity" factor includes the quality of staffing and their skills in managing the projects well, but it has a broader meaning. "Government capacity" also includes factors at the system level: whether the government system is structured or has mechanisms to support the smooth implementation of projects. If the government of the recipient country has a system that functions well for implementing projects or any other public services, then the implementation would go well (Interview #8[44]). Additionally, the government should ideally "have a strong vision of their future and map out their

[41] Interview #7 (May 5, 2014)
[42] Interview #9 (May 13, 2014)
[43] Interview #15 (May 1, 2014)
[44] Interview #8 (May 13, 2014)

strategy," and maintain "mechanisms for ongoing consultation and engagement throughout the life of the project."

On the other hand, when the interviewees mentioned that the capacity of the government is considered weak, they specifically meant the quality of local staff. Examples include lack of technical expertise (Interview #2, #6[45]), and no understanding of how to implement World Bank projects, including the World Bank procedures and procurement rules (Interview #9, #14[46]). Not having staff in place who would carry out the project is also an issue, as well as issues of staff not satisfactorily implementing the project, understand and owns the project (Interview #1[47]). Lack of commitment and incentives among the local staff could also be an issue (Interview #6[48]), as well as lack of experience in managing and implementing projects (Interview #9[49]).

Quality of staff

Staffing and skills include the quality of Bank staff and the local staff. I separate the Bank staff quality from that of the local staff. Quality of the Bank staff consists of six elements: ability to manage resources and teams, Bank staff being on the ground, cultural tolerance, experience, personality of the staff, and understanding education. Bank staff being on the ground means that the staff being really based in the country not on supervision missions. It is important that the staff visits the field on a daily basis because the staff, when they are on the field (as opposed to being in the office), can troubleshoot

[45] Interview #2 (May 14, 2014), Interview #6 (May 27, 2014)
[46] Interview #9 (May 13, 2014), Interview #14 (May 6, 2014)
[47] Interview #1 (May 12, 2014)
[48] Interview #6 (May 27, 2014)
[49] Interview #9 (May 13, 2014)

in real time and conduct real, physical monitoring (Interview #7[50]). Physical monitoring for construction of schools can be done by the staff going out to the field and see whether the school was built, interview the contractors and some of the construction workers on the ground.

Cultural tolerance is the Bank staff's patience and willingness to understand the other person's point of view and see through the cultural differences, the cultural sensitivity to appreciate whatever nuances (Interview #9[51]). Also, it is obvious that the more experienced the staff is, the more likely that a project will receive better evaluation (Interview #9[52]).

Personality of the individual staff is also important. Innovativeness, willingness to take risks and ability to adapt to changes in the project environment would depend on the personalities of the staff. A "good" personality of a task team leader that does a great job of implementing and supporting the government includes having an honest conversation and recognizing problems or issues and discussing how to solve the issue (Interview #5[53]). The personality of the task team leader not only affects the overall project performance but also the quality of the relationship between the Bank and the government. This will be further discussed in the later section.

Lastly, understanding education is another crucial component of the Bank staff. This becomes particularly important if the project is about improving the quality of learning. Staff understanding on classrooms, how learning happens and the dynamics of

[50] Interview #7 (May 5, 2014)
[51] Interview #9 (May 13, 2014)
[52] Interview #9 (May 13, 2014)
[53] Interview #5 (March 12, 2014)

learning would enhance the quality of the design and the implementation of education projects (Interview #12[54]).

The quality of local staff also matters to project outcomes. At the project level, the country would need a strong project leader who has political and technical skills that enable him/her to assemble a team and implement the project as designed (Interview #1[55]). For a well-implemented project, the local staff should not only have expertise in the area, but also know about the World Bank procedures to implement the project, have commitment and incentives to spend time on the project, be energetic and motivated.

Staff continuity

Frequent change of local staff and task team leaders

Middleton's analytical framework consists of staff continuity as part of management system factors. The interview data also supports the fact that staff continuity is can be correlated with project outcomes. Since each project lasts at least five to ten years, change of key personnel who are in charge of implementing the projects is inevitable. The staff change can occur from both the Bank and the government side, however, interviewees mentioned that there is more change of staff at the government side rather than the Bank side (Interview #7[56]).

There could be numerous reasons for frequent change of staff at the government side, but change in the political environment seemed to be the biggest factor. The change of the governing party will bring about changes in the structure of the ministry and

[54] Interview #12 (May 6, 2014)
[55] Interview #1 (May 12, 2014)
[56] Interview #7 (May 5, 2014)

personnel, particularly at the high official level. The change in the government delays the implementation of the project until all the high and mid-level position officers are in place. Thus, the delay in implementation may lead to a smaller number of project objectives being achieved, which will result in lower project outcomes (Interview #2, #7[57]). Delay in implementation could last up to a couple of months.

The interviewees also acknowledged that frequent change of task team leaders from the Bank might also lead to lower project outcomes. The task team leaders change more frequently in countries where there are difficult political and social situations (Interview #3[58]). The political and social situations are also part of the project environment factors, which is related to project performance. Thus, careful interpretation is needed for staff continuity on project outcomes since the change of staff is related to other project environment factors and operational management factors (e.g. delay in implementation) as well.

Theme 3: Ownership by the country (project environment factor)

A new theme (category) emerged from the interview data as part of project environment factor. I named it "ownership by the country." To clarify the term, the "ownership" here refers to a psychological ownership of the project by the government. Legally, the borrower owns all the projects. The importance of ownership of the country for successful implementation of the project has been mentioned by 13 interviewees. This is the largest number of sources among all coded categories. The number of sources for theme 1 "project design" and theme 2 "government capacity and staffing skills" was nine

[57] Interview #2 (May 14, 2014), Interview #7 (May 5, 2014)
[58] Interview #3 (April 21, 2014)

each. This implies the significance of ownership of the country may not be any less than project design variables or government capacity and staffing skills.

I define "ownership of the country" as the leadership of the recipient government in implementing the project, their commitment to the project, giving priority to implementing the project successfully, and having a vision for their education policy. "Ownership" here applies to both the central and local government. More specifically, a recipient government with leadership will have charismatic people at the high official level, who would prioritize the education sector and know exactly what they want to establish. Government commitment and leadership at the high level is extremely important as in developing countries, the technical staff is afraid to take decisions unless they are supported by their bosses (Interview #4[59]). This is regardless of whether the government is democratic or autocratic. An autocratic government like Rwanda also has a strong leadership and when they announce their plan, they actually make the plan happen (Interview #12, #13[60]). Thus, a government that has strong leadership also has an ownership of implementing the education project because they have strong commitment to carry out the project. However, this finding needs to be interpreted with caution, as it is a complex factor. Government leadership depends crucially on the political and economic context in the country at the time. Political economy-type of constraints are often the most challenging barriers in effectively exercising political leadership, especially in slow-growing or stagnating economies.

A government that has ownership of the project also prioritizes education sector rather than putting off educational agenda until after other sector agenda is settled. One of

[59] Interview #4 (May 5, 2014)
[60] Interview #12 (May 6, 2014), Interview #13 (May 7, 2014)

the interviewees mentioned difficulties in carrying out an education project in a country where there is a sentiment that education is not a priority for them.

> I worked in Laos, it was very different from Vietnam and much of the rest of Asia in that, the government just simply at that time did not think that education was important. The leaders of the countries, some of them didn't even have a secondary education, much less a college education. So education was not something that was important to them. So, the position of education in the society, makes a great deal of difference (Interview #9[61])

If education sector is not a priority for the government, there will be frequent delays in implementing an education project, which will lead to lower ICR/IEG ratings.

Not only prioritizing the education sector but also prioritizing the project itself at the highest level (usually the minister) can affect the overall process and speed of implementation of the project. As there is a bureaucracy in every country, the amount of time spent in implementation of the project depends very much on how fast the high-level officials approve and agree on the details of activities and procurement of the project. When the decision-maker at the highest level (the minister) does not prioritize the project, the procurement and disbursement of funds get delayed, making it difficult to implement activities within the planned time frame. If the activities are not conducted, it is very likely the project will get low ratings (Interview #14[62]).

A government that has ownership also has a vision of where they want to be in the future and have a strategy of how to get there (Interview #1, #6[63]). An interviewee pointed out China as an example of a country where the strong ownership of the

---

[61] Interview #9 (May 13, 2014)

[62] Interview #14 (May 6, 2014)

[63] Interview #1 (May 12, 2014), Interview #6 (May 27, 2014)

government significantly contributed to the successful implementation of the project (Interview #1[64]).

> I would say the Chinese government and those with whom you work with the government, have a very strong sense of what's happening around the world. They understand what other countries are doing. And they, I think, fairly well understand their own problems. As a result of that, they have a pretty strong vision of what they want to do and how they want to go about it (Interview #1[65]).

Governments would have ownership of the project by having a long-term vision of the future of the country and mapping out the strategy. This is also related to the government's capacity to create their vision and strategy. Therefore, government ownership is in part affected by the management system factors (government capacity).

The ownership discussed by the interviewees is not only at the commitment and prioritizing the project at the ministry level but also (the commitment and prioritizing) at the local level. Ownership at the highest level of the government is important because the bureaucratic process needs to be carried out before any disbursement of funds. Ownership at the local (or bottom) level is also necessary since this is where implementation takes place (Interview #1[66]). A project would not proceed with the commitment and ownership at the high official level, if there is lack of commitment, lack of incentives, and lack of expertise at the local level (Interview #6[67]). In sum, the government ownership includes a vision of the country, commitment, leadership, and prioritizing the education sector and the project at the high official level, and engaging

[64] Interview #1 (May 12, 2014)
[65] Interview #1 (May 12, 2014)
[66] Interview #1 (May 12, 2014)
[67] Interview #6 (May 27, 2014)

local officials to make commitments for the project. Several other factors were found to be important for receiving high evaluation ratings.

Relationship between the Bank and the government (operational management factor)

The interaction between the World Bank and the government as part of the operational management factor turned out to be one of the most significant factors for project performance. Government here refers to many different forms: project implementing agency, Ministry of Education, or Ministry of Finance. The term government in this study is defined as "the agencies that implement projects", usually the borrower government (mostly Ministry of Education) themselves. If the implementing agency is not the borrower government, in that case it is the government (whichever ministry) that outsource the project to the local agency. Of the 15 interviewees, 12 remarked about the relationship between the Bank and the government and explained how the relationship affects the implementation of the projects and thus the overall evaluation ratings. The number of interviewees who mentioned the relationship between the Bank and the government (12) is the second largest number of sources (interviewees) after the category "ownership of the country." Though the frequency chart shows that the number of times the word is mentioned in the data is less than half of "ownership of the country," the fact that it is mentioned by various sources suggests the importance of the category.

The frequent interaction (such as email exchange, and keeping the administrative staff from both sides up to date), continued dialogue, building trust by having a long term

relationship, sharing opinions on policy, sharing information on the project, being able to put issues on the table, and focusing on the same objective of the project are the characteristics of a "good" relationship between the Bank and the government. Frequent communication is essential in maintaining a good relationship between both parties.

> I think we probably talk by phone once a week, but certainly more than that. Just had to constantly send emails to each other. And I would do that also with the ministry of finance and we had a fiduciary agent that helped administer the project that I would be constantly in contact with them as well. So it's important to do that (Interview #3[68]).

One interviewee commented that if the Bank and the government continue to have dialogue on the project, then it is likely that the Bank continues to help them in whichever way when they had problems, even if the problem is not part of the project (Interview #4[69]). The relationship also needs to be a long-term relationship, particularly given that both the government and the Bank need to build trust (Interview #6, #7[70]). Countries do not enjoy having to greet new people all the time (Interview #6[71]) and the Bank also needs time to figure out the new local officials if there is any structural change within the implementing agencies or ministries.

Some interviewees gave a more sophisticated definition of the relationship between the Bank and the government. One interviewee mentioned that it has to be a "meaningful" relationship where the government or the Bank staff is able to put issues on

[68] Interview #3 (April 21, 2014)
[69] Interview #4 (May 5, 2014)
[70] Interview #6 (May 27, 2014), Interview #7 (May 5, 2014)
[71] Interview #6 (May 27, 2014)

the table and share their opinions on the issue (Interview #11, #12[72]). More sharing of information on the project and honestly discussing on any issues associated with the project will lead to better quality of implementation. Another interviewee noted that the Bank does not try to take the superior position than the government by simply giving financial assistance and relying on the country for implementation. Instead, the Bank tries to maintain an equal relationship with the government and ensure that they share information and opinions on policies with the countries, and also take the examples of best practices to other countries (Interview #11[73]). Thus, to maintain a "good" relationship between the Bank and the government, it is necessary that both sides be true partners by sharing information and be able to share opinions on any issues related to the implementation of the project.

The personalities of the staff are associated with the quality of the relationship between the Bank and the government. One interviewee noted that he/she would have different relationships with different task team leaders and the dynamics of these relationships somehow affect the project, or at least partially affect the project preparation or implementation (Interview #2[74]). Different relationship would be formed based on who the Bank and local staff is, and thus it would affect the quality of implementation differently. It should be recognized that the quality of Bank or local staff as part of the management system factors could influence the interaction between the borrower and the Bank, which is part of the operational management factors.

[72] Interview #11 (March 11, 2014), Interview #12 (May 6, 2014)
[73] Interview #11 (March 11, 2014)
[74] Interview #2 (May 14, 2014)

One should make careful conclusions based on this finding. It is clear that a good relationship between the Bank and the government would at least partially affect the quality of implementation of projects. Yet maintaining a good relationship could be the underlying purpose of the project. Some interviewees have alluded that sometimes, the project is implemented, knowing that there are challenges in the implementation of the project, for the sake of maintaining good relationship with the Bank or with the country (Interview #2, #4[75]). Therefore, whether there will be a strong association between a "good" relationship and the project performance ratings is not certain. Based on explicit evidence from the interviews (Interview #2,#3[76]), I conclude that a "good" relationship between the Bank and the government would at least partially affect the quality of the implementation of the projects.

Political environment (project environment factors)

Political environment is another factor that may significantly affect project outcomes. Political environment or context in this study is described as "policy stability" in Middleton's (1985) analytical framework. Political environment generally describes the overall political, economic, and social stability or change occurred during the project period that may influence project implementation and outcomes. The interviewees particularly noted the importance of project outcomes affected by the changes of the government due to changing political context (Interview #2, #6, #10, #13, #14, #15[77]).

---

[75] Interview #2 (May 14, 2014), Interview #4 (May 5, 2014)

[76] Interview #2 (May 14, 2014), Interview #3 (April 21, 2014)

[77] Interview #2 (May 14, 2014), Interview #6 (May 27, 2014), Interview #10 (May 6, 2014), Interview #13 (May 7, 2014), Interview #14 (May 6, 2014), Interview #15 (May 1, 2014)

When a different government comes into power during a project implementation, the structure of the ministries changes and new staff comes in. Then it is highly likely that there will be a delay in implementation of the project and the relationship between the World Bank and the government changes (Interview #7, #13[78]). Also, it would be very difficult to sustain long-term programs (Interview #6[79]).

Two interviewees commented on the delay of implementing projects. One interviewee commented on the delay at the project preparation stage due to changes in the government structure. The changes in the internal government, including the replacement of minister, vice minister and other high level officials, led to long delays that took about six months to move forward with a project preparation, agreement and negotiations that required high level decision making and bureaucracy (Interview #2[80]). Another interviewee commented on the delay on the procurement that requires approvals from the officials due to structural change (Interview #14[81]).

In sum, the data shows that the political environment does not directly affect the project outcomes itself, but through other factors at the management system and operational management level. The changing political context that leads to changes in the structure of the government and incoming of the new staff leads to changes in the relationship with the World Bank as well as delays in implementing projects. The staffing factor at the management system level, the input management and relationship between the Bank and government at the operational management level are affected by the political environment and changes of the government.

---

78 Interview #7 (May 5, 2014), Interview #13 (May 7, 2014)
79 Interview #6 (May 27, 2014)
80 Interview #2 (May 14, 2014)
81 Interview #14 (May 6, 2014)

Input management (operational management factor)

The interviewees mentioned some issues related to delivery of project outputs including disbursement and procurement while discussing factors for project outcomes. Most interviewees noted delays as disbursement and procurement issues that are likely to result in lower project evaluation ratings. On the other hand, if there is no delay in disbursement, the inputs will be delivered on time thus making it easier to reach the targets (Interview #11[82]).

There are numerous reasons for delays in procurement. These reasons are also factors that are listed in the project environment and management system factors. Delays of implementation could be due to political issues (Interview #1, #7[83]). Countries that are under political conflict may close their office in the worst case, thus implementing projects under such circumstances is not possible (Interview #7[84]). Another project environment factor that is related to delays in implementation is changes in government structure and bureaucracy, as mentioned in the previous section in detail.

Delays in disbursement can also be due to misprocurement. A number of interviewees noted misprocurement due to the rule of the Bank not being followed strictly mainly because the local staff is not familiar with the World Bank procedures (Interview #1, #9, #11, #14, #15[85]). Miscommunication between lawyers at the Bank and the local government side can also lead to delays in implementation of projects (Interview #2[86]). Corruption could be another factor for delaying procurement. One

82 Interview #11 (March 11, 2014)

83 Interview #1 (May 12, 2014), Interview #7 (May 5, 2014)

84 Interview #7 (May 5, 2014)

85 Interview #1 (May 12, 2014), Interview #9 (May 13, 2014), Interview #11 (March 11, 2014), Interview #14 (May 6, 2014), Interview #15 (May 1, 2014)

86 Interview #2 (May 14, 2014)

interview mentioned that in a school building project, the local contractor disappeared, which led to delays in implementation of the project (Interview #4[87]). A project not being the priority will also delay disbursement, which lead to objectives not being achieved and thus low ratings of the project (Interview #14[88]). This implies that government ownership factor discussed above is influencing the input management factor as well as project outcomes.

Bank support (Institutional building)

Although the interviewees noted that the Bank's role is not to implement the projects and the implementation is entirely upon the government, many of them have acknowledged the need for the Bank's support during the implementation. The "Bank support" factor overlaps much with the "institutional building" factor which will be discussed in the next section on validity of the ratings. However, it should be noted that the "institutional building" is a concept much more goal-oriented than "Bank support." The term "Bank support" is used as a factor that influence project outcomes, whereas "institutional building" used in the latter section indicates a factor that potentially raises question on the validity of the evaluation ratings.

Based on the data, I defined the Bank involvement, or support for the project as "institutional building (capacity building), continued dialogue and consultation on project or policy, the Bank staff being on the ground and providing technical support for the government, and conducting supervision." Capacity building or institutional building is definitely one form of Bank's support for the country. The Bank staff provides on-site

[87] Interview #4 (May 5, 2014)
[88] Interview #14 (May 6, 2014)

help for the government during the implementation of the project, so that in the long-term the government can build capacity to implement similar projects on their own.

> In the end, we're not supposed to implement projects, we're supposed to supervise them. But we're basically doing a lot of things that the government should be doing. But in order to make things move, we're doing them for them and with them. And in order to build a capacity (Interview #7[89])

The Bank should also provide meaningful and continued dialogue through an established mechanism and be engaged throughout the life of the project (Interview #4, #12[90]). One interviewee emphasized that the Bank staff being on the ground makes a huge difference to project outcomes. Nonetheless, the staff could communicate through phone calls or other internet technology, be available in the country and be engaged in the project the whole period so that they could troubleshoot on a real time basis, which makes a large difference in project performance (Interview #7[91]). Providing technical support to the government at the right time is important, thus sometimes the Bank staff would sit in the ministry for 15–20 days and if the government struggles with something, the Bank provides assistance to them (Interview #4[92]).

Corruption

Corruption was also mentioned frequently by the interviewees (7 interviewees), particularly when they talked about factors for failures of projects. According to the interviewees, corruption occurs both at the project implementation process as well as at

[89] Interview #7 (May 5, 2014)
[90] Interview #4 (May 5, 2014), Interview #12 (May 6, 2014)
[91] Interview #7 (May 5, 2014)
[92] Interview #4 (May 5, 2014)

the different levels of staff. More specifically, corruption means leakage in resources and often occurs in procurement and financial management (Interview #4, #15[93]). Money is not being used as it was committed to being used, and thus projects can get suspended (Interview #1[94]). Corruption may occur by contractors or local officials (Interview #4, #6[95]). Local officials would siphon off the materials of classrooms and school uniforms and make the local people pay for that when they are meant to be free (Interview #6[96]). Another example is that for a project that involves school construction, when there is corruption, the project never starts (Interview #10[97]).

There are certain types of projects in which corruption can occur more frequently. One interviewee mentioned that construction and textbook projects can be more corrupt than other types of projects (Interview #12[98]). Also, some countries are well known for corruption and the Bank staff are well aware of it and be prepared for possible corruption when they start the project (Interview #1[99]). The Bank staff can expect any potential corruption that may happen during the implementation of the project and be prepared to respond to it. Usually the Bank is informed of corruption from the third party, which is an outside source (Interview #1[100]). The Bank's response when they are aware if any corruption is happening with the ongoing project is help building a national system to detect problems and act on them, which is part of what is called "institutional

[93] Interview #4 (May 5, 2014), Interview #15 (May 1, 2014)
[94] Interview #1 (May 12, 2014)
[95] Interview #4 (May 5, 2014), Interview #6 (May 27, 2014)
[96] Interview #6 (May 27, 2014)
[97] Interview #10 (May 6, 2014)
[98] Interview #12 (May 6, 2014)
[99] Interview #1 (May 12, 2014)
[100] Interview #1 (May 12, 2014)

development." (Interview #6[101]). Moreover, the Bank established a department within the Bank that is devoted to detecting it and fighting for it (Interview #11[102]).

When the interviewees were asked if there is any action taken when corruption is going on during implementation, some of them answered that they would stop the project when the Bank is informed of corruption; however, since corruption is difficult to detect directly, they would simply reduce the number of activities (Interview 4[103]). Also, when evaluating the overall project from an ICR, the interviewees mentioned corruption is difficult to be detected from an ICR. Thus, the project performance rating at the World Bank may not include a potential corruption, leaving a question for the validity of the project performance rating.

Validity of the ratings as an emerging theme

The interview data not only showed evidence for factors that mattered to project outcomes, but it also illustrated some evidence that the current World Bank evaluation ratings system may not be measuring the overall project outcomes accurately, particularly the long-term outcomes. Although short-term outputs and long-term outcomes are distinguished in this study, there should not be a large discrepancy between short-term outputs and long-term outcomes, as the short-term outputs should indeed be directed towards long-term outcomes. However, the interview data in this study shows that some factors meant to contribute to the short-term output (IEG ratings) are, in fact, not the

[101] Interview #6 (May 27, 2014)
[102] Interview #11 (March 11, 2014)
[103] Interview #4 (May 5, 2014)

factors that should determine the success of long-term outcomes (impact in the community, sustainable outcomes of the project).

Focused target/design

While focused activities and narrowing the target population may lead to better project performance results, these may not necessarily be the factors for a "successful" project that has a long-term and sustainable impact on the local community. One interviewee mentioned that although the task team leaders are being pressured to simplify as evidenced by recommendations in IEG reports, education needs to be addressed cohesively and thus a project should be implemented sector-wide (Interview #12[104]).

> Obviously when you talk to the task team leaders, from their side the pressure is to streamline, to simplify……and when you can look in 50% of IEG reports you'll find some version of the recommendation, that this project was too complex, it should have been simplified…And there's nothing wrong with focus, and there's nothing wrong with having a good targeting. But education is a messy complex business…and if you're going to provide a sector cohesiveness in your support, then just picking a one piece, say we World Bank are only going to look at teacher training, and DFID is only going to look at textbooks, that doesn't work either. All the donors have their own timing issues and delays and setbacks, you've got to be in at least a cohesive dialogue that looks at the whole sector. (Interview #12[105])

Another interviewee also mentioned that the staff is being pressured in two different directions. One side is that the staff recommends not making the project complex and suggests keeping the objectives simple. The other side is that staff is being

[104] Interview #12 (May 6, 2014)
[105] Interview #12 (May 6, 2014)

praised for taking a complex project to the Board (Interview #14[106]). Thus, "focused target" may be an apparent factor for receiving higher project ratings, but it may not be a good strategy for a sector-wide reforms in education. Sector-wide reforms may result in the long-term success of the project. Therefore, we can conclude that simplified and focused target and project objectives may not necessarily be associated with the long-term project success.

However, four interviewees made a cautious interpretation of an ambitious project being a "failure." An ambitious project would set the target indicators of the project ambitiously and the IEG basically punishes a project for being ambitious (Interview #7[107]). Since the difficulty of the design is not considered in rating the project (Interview #8), if staff set a very ambitious targets, it is highly likely that the project will not achieve the goals and then it is a failure for the wrong reasons (Interview #11[108]).

Several interviewees emphasized more on designing ambitious projects than other interviewees. One of the interviewees reflected his/her past experience when a project set out high target indicators and thus being ambitious in terms of the project goals, however, did not reach the target indicator by the end of the project (Interview #12[109]). This project is likely to achieve lower evaluation ratings. However, the interviewee pointed out other aspects of project success: that they were able to reach policy agreements while doing the project (and thus building trust with the borrower) when the country was having economic crisis and various internal conflicts (Interview #12[110]). Also, the girls'

[106] Interview #14 (May 6, 2014)
[107] Interview #7 (May 5, 2014)
[108] Interview #11 (March 11, 2014)
[109] Interview #12 (May 6, 2014)
[110] Interview #12 (May 6, 2014)

enrollment became almost equal to that of boys when the population of the country was growing fast.

> I always believed in very ambitious project design. I believe that you need to be ambitious in design…..And you can't say we're just going to put kids in school and do nothing about quality. So you have to pay attention to quality. You can't say we're going to buy books and not train teachers. You have to have a complex set of factors coming together here….cause a lot of colleagues, junior colleagues would shy away from these things, because of these evaluations….but then you're going to design a project that leaves out all the risky things…Going from 30% of kids in school over this period of about ten years to having 60 or 70. I think we went for 60%. And we ended up getting 53% instead of 60%. So we were evaluated as not having been successful (Interview #12[111])

Another interviewee emphasized long-term sustainability of the project objectives rather than looking at short-term outcomes.

> But if you have more complex objectives, such as setting up an agency that is responsible for textbook contracting and quality control and distribution, well if you have that in, are you successful in getting that in place or not, and having it function is a hard thing to measure. But in my opinion that's more important. Because I think we're in the job not just creating outputs and even outcomes or results, we'll be creating things on an ongoing basis (Interview #15[112])

Therefore, focused target and less ambitious project design should be distinguished from factors/conditions that are associated with the "success" of projects or projects that have a significant and sustainable impact on the local community. An example of a successful education project is to pursue 100 percent children who are in the official school age to enroll in schools and ensure the children are actually learning in their schools. The project should set up its objectives that include increasing children's

[111] Interview #12 (May 6, 2014)
[112] Interview #15 (May 1, 2014)

school enrollment and improving the teacher quality. If one wants to receive higher ratings, one could only set up the project objectives as increasing children's school enrollment. A project with simplified goals and less activities would be much more likely to perform better. However, having both objectives on access to education and improving the quality of education, although challenging, may be necessary to improve the borrower country's education. Thus, a complex project with many objectives and activities would be more likely to show "success" and long-term impact of a project.

The above data shows that factors for high evaluation ratings may be different from the factors for long-term impact and success of the project. The data also raises the question on the validity of evaluation ratings. From information, it seems clear that ICR/IEG ratings do not fully represent the project success or the project's long-term sustainability. Thus, researchers need to further explore the validity of ICR/IEG ratings and the extent that these ratings represent the actual project success and long-term educational development in the borrowing countries.

Institutional building (operational management factor)

"Institutional building" is a term frequently used by the interviewees. Without this term, one would not be able to fully understand the nature of World Bank projects. Institutional building means providing technical support, increasing the government's capacity to implement projects, and making a sustainable system or structure within the government to carry out projects. The interviewees have mentioned various activities as examples of institutional building, including the Bank's support for sustainable judicial, legal system within the government, providing technical support, building capacity of the

government and building a relationship between the Bank and the government (Interview #3, 4, 6, 7, 10, 11, 12, and 14[113]). It can also be used as political terms, a term used to help the government address the issue of corruption (Interview #6[114]).

The most commonly used definition of "institutional building" by the Bank staff is an activity that essentially lay the foundation and investing for capacity building of the government (Interview #7, #10[115]).

> Many times you're looking at whether you built the capacity inside the ministry that you're dealing with, within the government itself, have you aided them to engage more with local NGOs and civil society….you want to see if the ministry is increasing its capacity to do what its own plans are…. (Interview #14[116])

More specifically, institution building includes putting some basic rules into place such as strengthening judicial and legal system and leaving a sustainable structure within the ministry (Interview #3, #6[117]). Thus, the role of the Bank to support in the long term is essential (Interview #6[118]).

One of the interviewees emphasized that although the term institutional building covers many different meanings, including the Bank providing long-term support to build government capacity, it covers the term political as well (Interview #6[119]). According to the interviewee, under Wolfensohn, the term "institutional building" was used as a code word for political development, and one of the main areas of institutional development

---

113 Interview #3 (April 21, 2014), Interview #4 (May 5, 2014), Interview #6 (May 27, 2014), Interview #7 (May 5, 2014), Interview #10 (May 6, 2014), Interview #11 (March 11, 2014), Interview #12 (May 6, 2014), Interview #14 (May 6, 2014)

114 Interview #6 (May 27, 2014)

115 Interview #7 (May 5, 2014), Interview #10 (May 6, 2014)

116 Interview #14 (May 6, 2014)

117 Interview #3 (April 21, 2014), Interview #6 (May 27, 2014)

118 Interview #6 (May 27, 2014)

119 Interview #6 (May 27, 2014)

was trying to help the government get rid of corruption. I conclude that the term "institutional building" can vary across different countries depending on the local context.

The term "institutional building" was mentioned by the interviewees as an inexplicit goal not directly noted in the ICR. That is, sometimes the project is implemented in order to have a good relationship with the borrower country. The relationship built by implementing a project helps the government to make major policy reforms with the support from the World Bank.

> It's more on the relationship. Part of what we do the project becomes kind of an excuse for having a relationship. So there's a lot of discussion about policy overall, reform overall, and government may be making major strides in the policy and reforms, that wouldn't be reflected in the project documents (Interview #11[120])

One interviewee indicated that regardless of whether the project is successfully completed, a project that has helped the government lay the foundation to build its capacity should be counted as part of the contribution to the government.

> 90% of what you do in the project is not financial. Most of the really important stuff is helping government get systems in place. Even if it didn't completely succeed, did you help set the ground work (Interview #12[121])

This implies that even though a project is evaluated as unsatisfactory in terms of achieving its objectives articulated in the ICR, the inexplicit goal of "institutional building" could have been achieved to some extent. Thus, a project that is considered to be a "failure" may be evaluated as a "successful" project when the degree of

[120] Interview #11 (March 11, 2014)
[121] Interview #12 (May 6, 2014)

"institutional building" is taken into account. Future evaluation indicators should need to reflect the inexplicit goal such as institutional building.

Validity of the ICR and IEG ratings directly mentioned

An emerging theme from the interview data is the validity of the ratings. Although most interviewees have described both positive and negative aspects of World Bank project performance ratings, a majority of interviewees mentioned concerns about the validity of the ICR and IEG ratings. They also discussed how these measures are not a perfect metric for project performance and suggested some alternatives for better measures. In this section, I discuss in detail about how and why the interviewees are worried about the validity of IEG as well as ICR ratings.

Negative aspect of project performance rating

ICR and IEG rating is mechanical

Several interviewees showed frustrations and limitations of ICR and IEG ratings. Particularly with ICR, a few of the interviewee has described that it is "very reductionist" and mechanistic ways to evaluate projects as the evaluation is more of "checking boxes" than actually trying to learn lessons from the project (Interview #9, #12[122]). Ticking off boxes appears to be an objective way to evaluate a project; however, it also means that the results indicators do not capture the context. Therefore, a lot of important stories do not get told as they are not what the Bank is asking about (Interview #9[123]). One interviewee said that particularly in case of Africa, a very significant improvement would

[122] Interview #9 (May 13, 2014), Interview #12 (May 6, 2014)
[123] Interview #9 (May 13, 2014)

never be guessed from looking at the project ratings (Interview #12[124]). Or projects that have not been successful can officially be considered as real success (Interview #12[125]).

Regarding the limits of IEG rating, one interviewee mentioned that an IEG evaluator may not have the right expertise to evaluate the projects since they need to evaluate many projects. Thus, the evaluation ends up becoming a more mechanized and routinized set of protocols that are applied in a superficial way and do not unpack or understand the processes underlying that the project succeeded or failed. Only the black and white distinction is present but not the nuance that gives us the lessons learned is absent (Interview #5[126]).

Context is not reflected in the ICRs

Many interviewees showed concerns about the problem of local context not being reflected in the ICRs. This is often also expressed as "nuanced elements" in the project. Some of them mentioned that they have seen unsatisfactory projects that would have received a better evaluation rating when nuanced elements of the project are considered (Interview #1[127]). For example, there is a great frustration with projects in Africa because Africa is well known as a difficult environment to implement projects. However, there is little appreciation for the context in the ICRs (Interview #9[128]). On the other hand, the reasons when a project shows positive result could be not due to the project but it could be simply the context that made the project seem successful. For example, if projects are

[124] Interview #12 (May 6, 2014)
[125] Interview #12 (May 6, 2014)
[126] Interview #5 (March 12, 2014)
[127] Interview #1 (May 12, 2014)
[128] Interview #9 (May 13, 2014)

implemented in countries that grow faster than other countries, the success of the project could be not due to the quality of the project but simply because the students are better fed and are able to do homework and other activities that are needed to achieve better outcomes (Interview #8[129]). Not only the ICR, but the IEG also does not take into account the context (Interview #7[130]).

Indicators do not measure the outcomes correctly

A majority of the interviewees have mentioned that the indicators that measure the outcomes are deficient and do not capture all the project outcomes (Interview #12, #7, #3, #4, #5, #1[131]). The project ratings are based on the extent to which the project objectives have been achieved. If all of the indicators are met, then the project is considered successful; thus ICRs rely too much on the results framework (Interview #3[132]). However, the definition of project objectives may be open to interpretation and the indicators do not measure the objectives precisely (Interview #7[133]). Also, there is a lack of understanding regarding what a good indicator should look like, and whether the objective is realistic (Interview #5[134]). Therefore, it is possible to conduct a project that receives highly satisfactory rating but is not capturing what needs to be done, meaning that a project can be overestimated on its performance (Interview #12[135]). Also, a project could have had some positive impact on the local community but can be rated as

[129] Interview #8 (May 13, 2014)
[130] Interview #7 (May 5, 2014)
[131] Interview #12 (May 6, 2014), Interview #7 (May 5, 2014), Interview #3 (April 21, 2014), Interview #4 (May 5, 2014), Interview #5 (March 12, 2014), Interview #1 (May 12, 2014)
[132] Interview #3 (April 21, 2014)
[133] Interview #7 (May 5, 2014)
[134] Interview #5 (March 12, 2014)
[135] Interview #12 (May 6, 2014)

unsatisfactory simply because the results framework indicators fail to capture the positive impact. In fact, many projects are not rated well because the indicators are not appropriately matched to the objectives (Interview #4[136]). One interviewee thinks that a lot of projects rise or fall under the choice of indicators (Interview #5[137]).

Setting the right indicators to measure project outcomes is particularly difficult in education sector, where there are positive and negative externalities, unexpected outcomes or spillovers (Interview #8, #11, #12, #3, #4[138]). The indicators may not be comprehensive enough to capture some of these externalities (Interview #1[139]). One typical positive externality frequently mentioned by the interviewees is capacity building and institutional building, which is much more valuable but very difficult to be quantified, therefore hard to be picked up in an evaluation (Interview #3, #7[140]). For example, an ICR would state that the relationship between the government and the Bank is good, but it is difficult to get a solid evidence for such statement (Interview #3[141]). One interviewee underlined the importance of process of implementing projects, as opposed to final results. According to him, a process that is dynamic and inclusive of local people is likely to lead the project into the right direction (Interview #12[142]). Sustainability (or long term benefits) is also an element that is difficult to be measured because only specific benefits are measured at the end of the project (Interview #4).

---

136 Interview #4 (May 5, 2014)
137 Interview #5 (March 12, 2014)
138 Interview #8 (May 13, 2014), Interview #11 (March 11, 2014), Interview #12 (May 6, 2014), Interview #3 (April 21, 2014), Interview #4 (May 5, 2014)
139 Interview #1 (May 12, 2014)
140 Interview #3 (April 21, 2014), Interview #7 (May 5, 2014)
141 Interview #3 (April 21, 2014)
142 Interview #12 (May 6, 2014)

Although many interviewees mentioned the appropriateness of indicators to measure achievement of project objectives, a few of them said that IEG evaluates the appropriateness of indicators for outcomes in ICR. One interviewee has mentioned that IEG examines whether the indicators were selected appropriately to measure what the project objectives were when they evaluate ICRs (Interview #14[143]). The IEG also observes whether real evidence was presented to support the fact that the objectives were met and without the project, the objectives were not met (Interview #3[144]).

Bias in ICR

Some interviewees have alluded to the possibility of bias in ICRs. The possible reason for such bias could be some personal attachment to the project or fear of disconnect between the ICR rating and the IEG rating. However, the direction of bias can be different according to the reason. If one is attached to a project, it is likely that one will overestimate the project outcomes. Although the ICR is written by a person not involved in the project, if a Bank's task manager helped with writing the ICRs, it is likely to be biased if the manager has some kind of ownership in the project or if they are going to work with the government in the near future. It is easy to be biased when the evaluator is partly involved in implementation of the project (Interview #6[145]). On the other hand, fear of disconnection between ICR and IEG may bias the ICR rating, but the bias is likely to be downwards. The staff tends to downgrade ICR ratings or at least be very cautious,

[143] Interview #14 (May 6, 2014)
[144] Interview #3 (April 21, 2014)
[145] Interview #6 (May 27, 2014)

because IEG asks for explanation on the disconnection between ICR and IEG ratings, rather than unsatisfactory project outcomes (Interview #9[146]).

> There is a sense of frustration that IEG always seems to be harsher than the Bank's own ratings, and I think as a result the Bank tends to be more conservative of how it rates its project. You don't want discrepancy with the rating that IEG gives (Interview #5[147]).

The Bank operational regions are monitored on their disconnect (Interview #8[148]). There is no direct relationship between staff salary and the ICR and IEG disconnect; however, managers are actually monitored on the rate of disconnect, thus they have incentives to make sure that the staff report the ICR ratings accurately (Interview #8[149]). The interviewees think that the fear of disconnect actually makes the staff more realistic as to self-assessment (Interview #8[150]). Therefore, the staff do not want to be too optimistic because they do not want their ICR rating downgraded (Interview #7[151]).

In addition to the possible bias in ICR rating, one interviewee indicated that the ratings could be very subjective. He had an experience when a project was evaluated to be moderately satisfactory but at the very borderline with moderately unsatisfactory, it had a lot of elements evaluated as moderately unsatisfactory. However, the project was finalized as receiving a moderately satisfactory and there were reasonable stories for the

[146] Interview #9 (May 13, 2014)
[147] Interview #5 (March 12, 2014)
[148] Interview #8 (May 13, 2014)
[149] Interview #8 (May 13, 2014)
[150] Interview #8 (May 13, 2014)
[151] Interview #7 (May 5, 2014)

result. There was no issue with the IEG regarding this evaluation; therefore, the interviewee thinks that the ratings can sometimes be very subjective (Interview #14[152]).

Positive aspect of the project performance rating

Although a majority of the interviewees showed concerns regarding the validity of both the ICR and the IEG ratings, quite a number of them mentioned positive aspects of these ratings at the same time. Several mentioned that IEG rating is useful, and is a good indicator and is "as objective as human can be" (Interview #1, #6, #9, #14[153]). The IEG rating is useful because it forces the Bank staff to design the projects carefully (Interview #11[154]). The ratings are "objective" in that the organization (IEG) is independent (Interview #10[155]). The ratings partially reflect reality, if not one hundred percent of all reality, as the rating depends on evidence and some evidence is difficult to achieve (Interview #3[156]).

With ICR rating, it can be an objective way to evaluate projects if one is only looking at numbers (Interview #9[157]). Also, the ICR writer attempts to be as objective as possible. The ICR writer is given the list of facts like memoirs of ISRs, interviews that he/she has done and other data that he/she collected and then the writer pulls them altogether and make a coherent story from the list of facts (Interview #14[158]). Therefore, the ICR rating can be as objective as humans can be. Also, one interviewee described that

[152] Interview #14 (May 6, 2014)
[153] Interview #1 (May 12, 2014), Interview #6 (May 27, 2014), Interview #9 (May 13, 2014), Interview #14 (May 6, 2014)
[154] Interview #11 (March 11, 2014)
[155] Interview #10 (May 6, 2014)
[156] Interview #3 (April 21, 2014)
[157] Interview #9 (May 13, 2014)
[158] Interview #14 (May 6, 2014)

the ICR ratings are definitely correlated with reality although they do not always reflect all of the reality (Interview #3[159]).

> I think that the ICR ratings are correlated with reality. You know, there's a high correlation. I don't know how high, but I mean it's not like it's random. There's certainly a relationship, it's just not a perfect relationship (Interview #3[160]).

Regarding the subjectivity and objectivity of the ICR and IEG ratings, interviewees acknowledge that the ratings adopt a framework that makes them as objective as possible, but there is still room for subjectivity depending on the ICR writer.

Another positive opinion on the ratings is that it is a good initiative of the World Bank as other organizations may not yet have such form of evaluation system. One interviewee thinks that overall the IEG evaluation of project performance is a good step, a good initiative and institutionally, a good activity (Interview #2[161]).

What alternative measure is needed besides IEG rating and what needs to be improved?

Sustainability of the project

Interviewees not only discussed the validity of the ICR and IEG ratings but also talked about alternative measures needed to improve the evaluation of project performances. A majority of interviewees pointed out that sustainability of the project, or long-term outcomes need to be considered in evaluating education projects (Interview #3,

[159] Interview #3 (April 21, 2014)
[160] Interview #3 (April 21, 2014)
[161] Interview #2 (May 14, 2014)

#4, #6, #8, #7, #11, #12, #13[162]). Sustainability usually refers to institutional development after the Bank withdraws the country. Some questions related to evaluating the sustainability of a project could be: did the organizations involved in the country get strengthened as a result of the project (value-added of the World Bank), and are the benefits of the projects likely to be sustained in the long run (Interview #3, #6[163]). It could also be if the project was sustainable in a sense that people believe in the project (Interview #3[164]). Although while on supervision, the Bank staff made sure that the government is considering sustainability (Interview #11[165]), such long-term benefits are difficult to measure (Interview #4[166]). Despite the difficulty to measure long-term benefits of projects, since other donor agencies include sustainability as one of the measures for project outcomes, measures for sustainability should be used in the Bank (Interview #8[167]). Sustainability is particularly important for education projects because the problems in the education sector tend to be continuous (Interview #13[168]).

Institutional building/capacity building

Institutional building and capacity building is another element emphasized by the interviewees to be included as ratings. The ratings do not reflect all the outcomes of the project, particularly in capacity building. Capacity building is an inherent objective, but

[162] Interview #3 (April 21, 2014), Interview #4 (May 5, 2014), Interview #6 (May 27, 2014), Interview #8 (May 13, 2014), Interview #7 (May 5, 2014), Interview #11 (March 11, 2014), Interview #12 (May 6, 2014), Interview #13 (May 7, 2014)
[163] Interview #3 (April 21, 2014), Interview #6 (May 27, 2014)
[164] Interview #3 (April 21, 2014)
[165] Interview #11 (March 11, 2014)
[166] Interview #4 (May 5, 2014)
[167] Interview #8 (May 13, 2014)
[168] Interview #13 (May 7, 2014)

often not stated explicitly (Interview #7[169]). Ideally, the Bank would take time to build the capacity within the ministry and ensure that the capacity at the implementation level exists so that later projects can be implemented smoothly (Interview #1[170]). However, often capacity building is not ensured (Interview #1[171]). Therefore, even if a project did not completely succeed, if the project helped set the ground work and helped with capacity building and institutional building, then the project should be acknowledged for its accomplishment (Interview #8, #12[172]).

Other elements to improve project evaluation measures

One interviewee mentioned that a better way to measure success of a project is to examine the actual results and take the context into account (Interview #7[173]). Another interviewee underlined that if one is interested in Millennium Development Goals (MDGs) and learning outcomes, the Bank should conduct more impact evaluations instead of using ratings (Interview #11[174]).

Dilemmas in the evaluation system

In thinking about the ways to improve the ratings, some issues in the process of giving ratings need to be dealt with. The first one is the dilemma of whether the Bank should support the countries that have difficult environment to get good rating. Certain countries are known for having an environment that is easy to get a good evaluation or

169 Interview #7 (May 5, 2014)
170 Interview #1 (May 12, 2014)
171 Interview #1 (May 12, 2014)
172 Interview #8 (May 13, 2014), Interview #12 (May 6, 2014)
173 Interview #7 (May 5, 2014)
174 Interview #11 (March 11, 2014)

very difficult to get a good evaluation rating. The Bank does not want to waste money in a country and expects a certain level of performance, but at the same time, the countries that do not have a conducive environment to perform well are those who need the Bank's support the most (Interview #12[175]). The staff is always judging on these issues and the evaluation metrics need to take that into account (Interview #12[176]).

The second issue in the process of rating is the dilemma that complex projects are something that the Bank compliments the staff for, but at the same time, the staff are encouraged to make the project objective simple. The projects that have simple objectives are much more likely to get better ratings. One interviewee said,

> We're always getting pushed in two directions. One, from the front office that looks at our projects' implementation and says you made this way too complex and you should make it simpler. And then the other side is, you get patted on the back, for taking a complex project to the Board. So it's always competing demands (Interview #14[177]).

These two issues indicate that the difficulty of implementing projects needs to be taken into account when projects are evaluated. The difficulty in the local context where projects are likely to have many issues during implementation, and also the difficulty in achieving project objectives needs to be considered.

The last issue in rating education projects is the method of combining different ratings and deciding on an overall rating. A project usually has several project development objectives (PDOs) and it is possible that part of the objectives were achieved while other objectives failed to reach their target. Then the overall rating

[175] Interview #12 (May 6, 2014)
[176] Interview #12 (May 6, 2014)
[177] Interview #14 (May 6, 2014)

becomes open to subjectivity. Regarding this issue, one interviewee pointed out that there is a need for the ability to differentiate the project objectives and not lump them in the entire rating (Interview #13[178]).

[178] Interview #13 (May 7, 2014)

CHAPTER VII

DISCUSSION

Generally, the findings from quantitative and qualitative methods show little overlap, but complement each other. While it is important to note that the little overlap between quantitative and qualitative data raises question on the validity of evaluation system, based on the main findings from quantitative data, the nuances added by the interviewees show that success of project implementation is quite complex and country specific. Both types of data indicate that country characteristics, to some extent, influence project outcomes. Country level variables in the quantitative data showed some statistically significant result, although the significance were sensitive to model specifications. The finding of importance of country characteristics, particularly government effectiveness, in this study aligns with Kaufmann et al.'s (Kaufmann & Wang, 1995) study that conducted subsample analysis of education projects and found that policy variables (black market premium, fiscal deficit) became more significant on project performance than a pooled sample of all projects.

On the other hand, project level variables in this study showed non-significant results for project outcomes. The interview data also illustrates the importance of country characteristics for good project outcomes. The success and failure of projects depends significantly on the local context and the political, social and economic environment of the country. However, the interview data also revealed the importance of project level factors (e.g. project design, quality of staff, staff continuity), thus, making it difficult to

conclude whether country characteristics matter more than project level factors for project outcomes. Unlike the finding from Denizer et al.'s (2013) study, whose main finding is that the success of individual projects varies much more within countries than it does between countries, this study shows that there is no evidence that country or project level characteristics are more important than the other. However, the interview data shows that there is complicated relationship between country and project level characteristics.

To answer the first research question, "what factors explain the variation in project performance of education development projects at the World Bank?," this study found that project performance depends largely on the local context and country characteristics, particularly on the government's effectiveness, commitment and leadership. The study, however, shows no evidence that project level factors matter less than the country characteristics. Instead, the interview data suggest that project design, project personnel, such as staff continuity, are significant elements for project performance, and each factor is associated with another factor, reflecting the complexity of project implementation. This finding answers my second research question "how do the Bank staff consider project determinants?" I discuss the implications of each of the determinants for project performance in this section.

The significance of the role of the government

One of the main findings of this study is that the role of the government is crucial for the success of World Bank education projects in developing countries. The quantitative data showed the significance of the government's ability in delivering public

services to its citizens (expressed as government effectiveness). Qualitative analysis indicated that the role of the government, particularly in terms of ownership, commitment and leadership were frequently mentioned, implying the importance of government responsibility in project success.

The importance on the role of government is related to the finding from interview data, that the relationship between the Bank and the government can greatly affect project outcomes. Governments that are politically more committed to project success tend to have better relationship with the Bank as the exchange of opinions and dialogue among experts may lead to better education sector strategy. A study that compared World Bank education loan in South Korea and Mexico illustrates a good example of how the relationship between the Bank and the borrower government can contribute to difference in overall national development (Lee, 2010). The South Korean government made use of World Bank monitoring and supervision team that composed of specialists during their visit to Korea by exchanging of opinions and advices over the policies and administration of education sector in general. Their active use of the experts from the World Bank contributed to the formulation and change of policies in education sector . On the other hand, Mexican government's relationship with the Bank, although cooperative, rarely developed into open-minded, professional discussions on the education sector issues and policies. Thus, South Korean sector analyses may have contributed to more efficient investment in education, that may have resulted in faster economic growth in Korea than in Mexico (Lee, 2010).

The finding that the effectiveness of borrower government on project outcomes also suggest that foreign aid can have an effect when the government has a high level of

effectiveness, meaning that citizens' perceptions of their government on the quality of public and civil services, the degree of independence from political pressures, the quality of policy formulation and implementation, and the credibility of the governments' commitment to such policies, are high (World Bank, 2015c). This finding is partly consistent with that of Denizer et al.'s (2013) study, who used the CPIA rating (country level rating for aid allocation) as a proxy for country quality. The CPIA rating emphasizes macro country-level measures of policy and institutional quality, therefore, to some extent similar to the measure of government effectiveness. Denizer et al. (2013) found that CPIA scores are significantly and positively associated with project outcomes across all sectors, including the education sector.

The analysis from the interview data supplements the finding on the importance of the government's role on project performance. The second theme discussed in the findings section indicated that government and institutional capacity emerged frequently in the data. Government effectiveness is interpreted as government institutional capacity, commitment and ownership in the interview data analysis. The government structure or mechanism to implement projects efficiently as well as staff that have technical expertise and capacity to implement projects are essential components of project success. More specifically, this means that in a government with the appropriate institutional capacity with little or no corruption in delivering projects to the beneficiaries, projects are implemented according to the project plan (or proposal), project is completed in a timely manner, and faces few issues in procurement and financial management.

This finding supports the Burnside and Dollar's argument that foreign aid will work in countries with sound policies. As explained in the literature review section,

Burnside and Dollar's (1997) study found that foreign aid had a positive impact on economic growth in developing countries with good policies, whereas with poor policies, aid had no positive effect on growth. Their argument of "sound policies" or "good policies" is referred to as economic policies such as fiscal, monetary, and trade policies. Also, their study examines the aggregate foreign aid at the national level, instead of project level aid. In my study, foreign aid is treated at the project level instead of aggregate aid (and limited to World Bank education projects only), outcome is project performance instead of economic growth, and "good policy" is interpreted as government effectiveness and commitment instead of fiscal, monetary and trade policy. Given these assumptions, my study supports Burnside and Dollar's (1997) findings that foreign aid has a positive impact on developing countries when the governments have "good" policies, particularly in terms of effectiveness, commitment and ownership to implement the foreign aid project.

Other studies conducted later also supported this main finding (e.g. Burnside and Dollar, 2000; Collier and Dollar, 2004; Dollar and Levin, 2005). Dollar & Levin's (2005) study, using World Bank project evaluation ratings data from 1990 to 1999, found that there is strong relationship between institutional quality and project success rate. Although they conducted the analysis at the country level and institutional quality is measured by the Rule of Law index, Freedom House democracy measures, the government effectiveness includes both these measures in addition to other measures such as corruption measures from Transparency International (Transparency International, 2015). Thus, the findings from my study indicate that the broader definition of government effectiveness, which includes more than rule of law and democracy measures,

matters significantly for project outcomes not only at the country level but also when the project outcome is measured at the project level.

My finding indicates that government effectiveness, what is known as a significant factor for project outcomes across all sectors, is also significant in education sector projects. Hassan (2012) found that governance is particularly important for education sector projects, although governance is important for most project sectors. My finding aligns with the findings of Hassan (2012). The difference between Hassan (2012) and my study is that Hassan (2012) used all six factors of Worldwide Governance Indicators, while I used the government effectiveness indicator only.

Within the larger international development community, this finding also reiterates the concept of "ownership" emphasized in the Paris Declaration on Aid Effectiveness (OECD, 2005). As explained in the research design section, Paris Declaration on Aid Effectiveness emphasized the role of government in implementation of development projects using a specific term "ownership." The finding from this study aligns with the latter part of the Paris Declaration on ownership, which states that the government need to play the leading role not only in developing development strategies, but also in coordinating foreign aid projects at all levels and maintain dialogue with donors. Combining the definition of "government effectiveness" as part of the Worldwide Governance Indicator, the finding in this study implies that the quality of government's public services to coordinate and deliver development projects in collaboration with donors is a significant determinant for education project outcomes. Thus, the finding of this study confirmed the emphasis on government ownership empirically.

However, caution is needed when interpreting the implications of this finding. The degree of government ownership to a particular project could be quite complex, especially when the government tries to introduce education reforms in the context of difficult political economy and/or economic decline. In fact, this was the case of SSA countries during most of 1970-2000 period. Compared to the East Asian countries, where reforms were introduced in the context of growing education budgets and declining school population, SSA countries had socio-economic challenges such as rapid growth in population of primary school age and stagnating budgets. Further, the stagnation of budgets in SSA countries during the 1980s and 1990s led many countries resorted to using contract teachers, however, insufficient attention by governments to the carrier development of these teachers have often created additional strains within the teaching force. This neglect reflects weak institutional capacity, including insufficient ability by governments to engage with teacher unions, particularly in SSA countries where the teacher unions are generally strong.

The fact that government effectiveness matters do not necessarily mean that the World Bank should support education projects in countries that have sound institutional capacity to implement the project. Rather, the countries with weak capacities need assistance the most, and the assistance should be given considering their political economy situations. In fact, the current shift of aid is focusing more on country needs rather than countries with strong performance. International development agencies are currently putting priorities to conflict affected and fragile countries. This study supports the current trend and policy, and recommends that more should be done for improving government effectiveness, raising their commitment in development projects, and in

doing so they should consider their political and economic environment. This study deals with the "classic dilemma", the trade-off between institution building and project effectiveness. The findings from this study implies that donors should help the borrower countries to build the institutional capacity for implementing projects successfully rather than increasing project effectiveness.

Inconclusive results of education spending on project performance

The reason for adding the education spending variable in the regression model was based on a hypothesis that if the government spends more on education, they value the education sector, therefore, this will favor the success of projects. However, I found a negative relationship between education spending and project performance. I have explained the reason in the results section, that education spending variable may be endogenous due to the country characteristics that spends more on education than other countries. I examined the descriptive statistics to examine education spending by region more closely.

Figure 20 shows that MNA region spends the most in education, Latin America and the Caribbean next, followed by Africa region. Since Table 17 shows that only four percent of all the projects are implemented in MNA region, I focus on LCR and AFR region. Looking at the Figure 21 that shows the average GDP growth by region between 1996 and 2011, it is clear that countries in Africa have grown economically during the period, while countries in the LCR were lagging behind countries in other regions. However, Figure 20 shows that LCR spend about the same share of GDP on education as

SSA but more than EAP and SAR region. Also, LCR is a region that has higher post-primary enrollment rates than other regions.

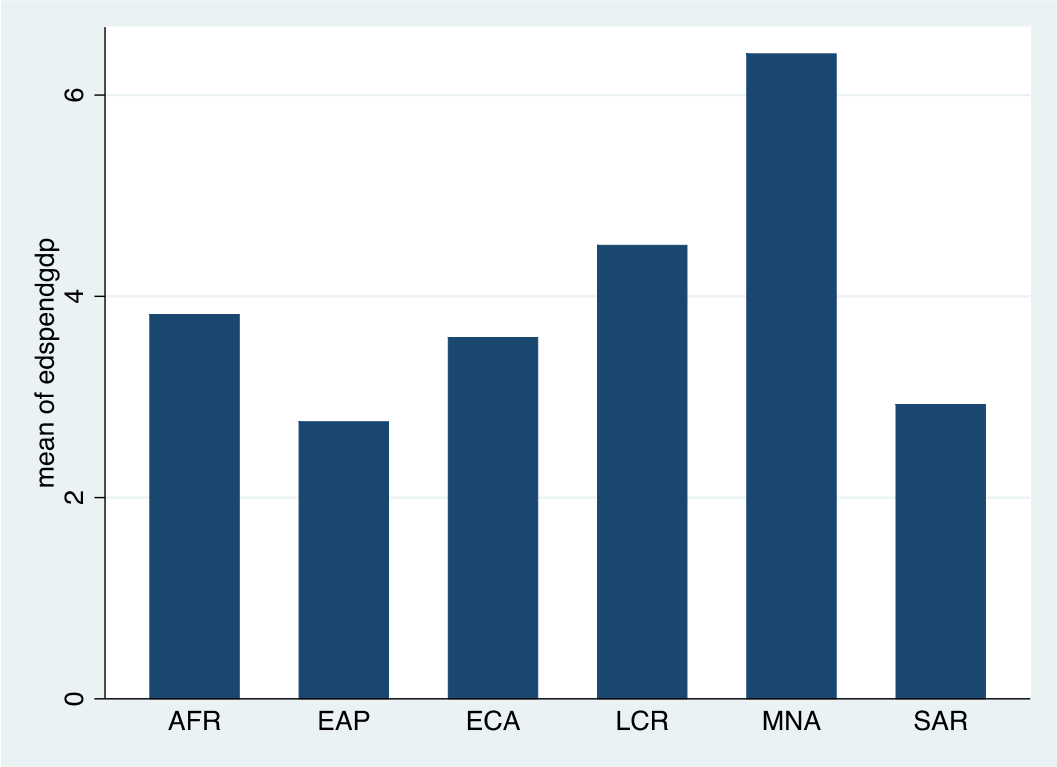


Figure 23 Education spending by region, 1996-2011

Table 17 Number of projects used in the regression analyses by region

Region	Number of projects	% of total projects
Sub Saharan Africa (AFR)	245	27.71%
East Asia and the Pacific (EAP)	89	10.07%
Europe and Central Asia (ECA)	84	9.5%
Latin America and the Caribbean (LCR)	304	34.49%
Middle East and North Africa (MNA)	41	4.64%
Southeast Asia (SAR)	121	13.69%

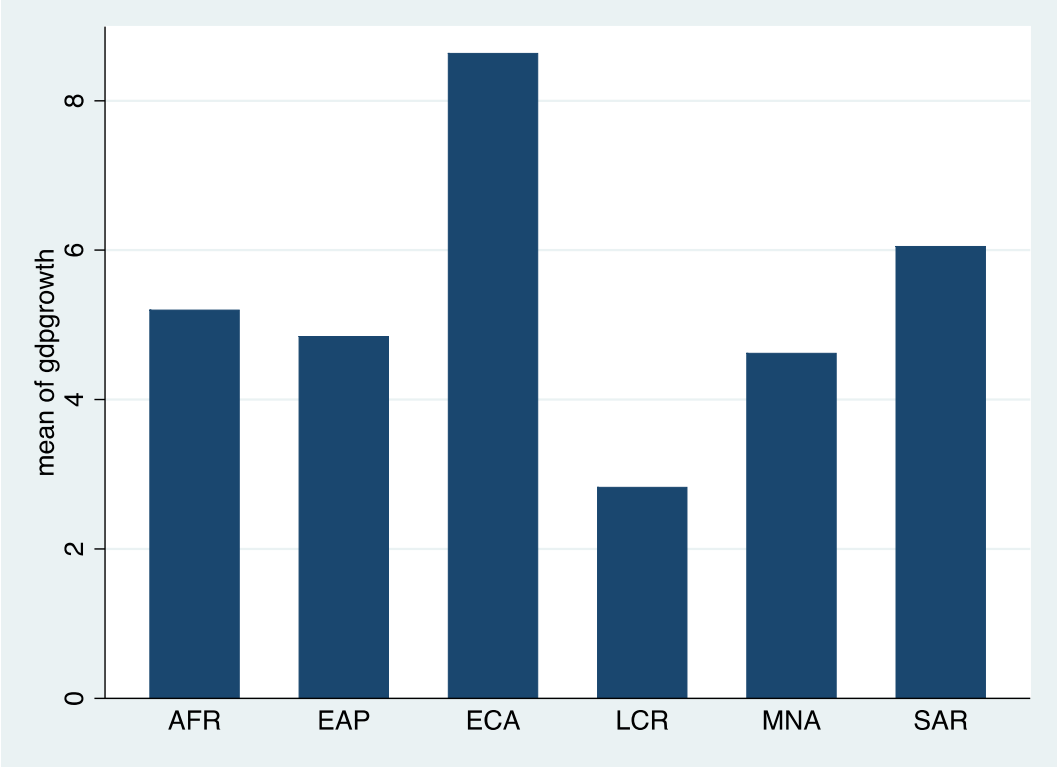


Figure 24 GDP growth by region, 1996-2011

However, the interpretation of these statistics on education spending in Latin America needs caution. Although the descriptive statistics show that the public expenditure on education as a percentage of GDP has risen in the LCR region, an OECD report analyzed that, on average, Latin America still spends about five time less per pupil than the typical OECD country, because the school-age population is much larger in Latin America than in OECD countries (Zoido, 2008). Moreover, more spending does not necessarily guarantee higher quality; and the report describes that Latin American countries do not appear to secure the greatest possible value from their expenditures (Zoido, 2008). In other words, if the same amount is spent in other regions, the outcomes are larger than the amount spent in Latin America region (Zoido, 2008). In addition, there is a large socio-economic gap in Latin America; thus, the spending for education for this

specific region should be focused on spending for lower income students rather than general spending in education.

In sum, the negative relationship between education spending as percent of GDP and project performance should be interpreted in the regional and local context. As an example, the negative relationship between education spending in Latin America and project outcomes may be due to the growing school-age population, inefficiency in education spending and achievement gap based on socio-economic status of Latin American students. The example of Latin America indicates that more spending would be important but how the money is spent matter much more. Therefore, although LCR region seems to spend more in education than any other regions from the graphs, if the countries do not spend the money to address their main challenge (e.g. socio-economic gap), more spending would not be directly related to better education project performance.

In case of Africa, the region has very difficult environment to implement projects due to political, economic and social instability as mentioned by various interviewees. With a growing GDP, Sub Saharan African countries' spending on education have also increased by 5% between 1999 to 2010, nevertheless, aid amounts as much as one-fifth of education budgets in poor countries, particularly in Sub Saharan Africa (UNESCO, 2012). Part of the reason may be the fact that overwhelming share of education spending in SSA is on teacher salaries, which is difficult to reduce. The poor countries that have increased spending on education still rely on foreign aid, which can be volatile or can be even stopped suddenly due to political instability in recipient countries or changing priorities in donor countries (UNESCO, 2012, p. 144). In such case, the countries that

spend more on education (but rely heavily on foreign aid) have a lower probability for education projects to receive high ratings because of the external factors related to the implementation of education projects. Another reason for this finding may be due to the difference in public funding for education between SSA and the rest of the developing world, which is accentuated by the large and growing differences in the growth of school-age population. Therefore, spending in education may be endogenous, making it difficult to conclude that spending more in education (and putting priority in education) is associated with good project ratings.

In short, the attempt to observe the relationship between valuing of education in a country (proxied by education spending) and project outcomes did not lead to any definitive conclusion due to potential endogeneity of the education spending variable. Since education spending was included in the model after preliminary interviews with some of the staff, the evidence from this study is not strong enough to confirm the determinants of project performance suggested by the interviewees. Future studies might need to include school age population variable in the analytical model, as we have seen in Latin American case, the growing school age population canceled out the effect of more spending on education.

East Asia and the Pacific region as a significant predictor for project performance

Both quantitative and qualitative interview data indicate that region is an important factor for project success, particularly for the projects implemented in the East Asia Pacific region. This is in contrast to Kilby's (2000) study, in which regional dummy for Asia did not show any significance for predicting the change in project performance.

Since Kilby's (2000) study examined projects across all sectors, the reason for a significant result of the East Asia regional dummy in this study could be due to the fact that it is specifically focused on education projects.

One of the reasons for a significantly larger probability of project success in the East Asia and the Pacific region may be cultural factors embedded in East Asia. One interviewee explained

> In East Asia, everybody did everything right anyway. This is comparable to other regions where there is no capacity to do evaluations like that (Interview #10[179]).

The underlying assumption for such statement could be the values and cultures in East Asia. East Asian values and cultures explain the success of Asian students in large-scale international student assessments (Leung, 2001), as well as a strong commitment to education within Asia (Kennedy & Lee, 2007). The highly value-laden nature of commitment in education may have led to a discrepancy between East Asia and other regions such as Southeast Asia, where high levels of illiteracy continue to exist (Kennedy & Lee, 2007). The value-laden commitment to education exists at all levels from family to government. Therefore, it is highly possible that the governments in East Asia have valued education projects and put more commitment and leadership than governments in other regions, leading to better project performance than other regions.

However, Figure 19 illustrated that government effectiveness is more important factor than whether the country is located in the East Asia Pacific region. Figure 19 implies that simply being an East Asian country and having value-laden commitment in

[179] Interview #10 (May 6, 2014) (Leung, 2001)

education will not always be associated with good project performance. Instead, the government of a country located in a non-EAP region but has high effectiveness will be more associated with good project performance.

Contrary to East Asia regional dummy, the Africa dummy variable did not turn out to be significant from the quantitative analysis in this study. This finding is consistent with the findings from Kaufmann et al.'s (1995) study, in which the Africa regional dummy variable did not show any significant results for predicting project performance for projects across all sectors. On the other hand, the qualitative interview data showed that projects implemented in Africa have more difficulties in successful project performance. Although this study cannot corroborate that projects implemented in Africa are less likely to show project success, future studies should continue to include Africa as dummy variable in the analysis as it is indicated as an important factor from the interview data.

Mixed finding on project level factors, still it is necessary to examine project design, the quality of implementation staff

The findings showed that the effects of project level factors on project performance are mixed. From the quantitative analysis, the results showed no evidence that any particular project level factors were significant determinant for project outcomes. Some of the hypotheses were that projects that have Education for All theme, Education for Knowledge Economy theme may have different outcomes, or projects implemented in primary education sector may be significantly different from those of the tertiary

education sector. These hypotheses were not supported by the findings from the quantitative data in this study[180].

However, the interview data indicated that some project level factors are important determinants for project outcomes. Interviewees have said that projects that have focused targets with less complicated objectives are more likely to achieve higher ratings. I have used two proxies for complexity of the project based on Denizer et al.'s (2013) study. One is the largest share of the project assigned to a subsector within the education sector. Higher values indicate less dispersion of the project across subsectors, and presumably also less complexity (Denizer, Kaufmann, & Kraay, 2013). Another variable used to measure the project complexity is whether the project is in the second or third phase (repeater project). However, neither of these variables (largest share of a project and dummy for repeater project) was statistically significant.

The inconsistent findings between the quantitative and qualitative data imply that hard data such as percentage of a project belongs to a specific subsector, theme, repeater, and the share of largest subsector is not sufficient to explain the probability of project success. Such an inconsistent finding also suggests that future studies need to examine the project level factors more closely, preferably the more detailed and nuanced factors and transform the qualitative components into hard data.

The project level factors that emerged as important determinant for success during the interviews were quality of staff, staff continuity, and project design. These factors are both mentioned by the theoretical framework suggested by Belassi and Tukel (1996) and Middleton (1985), as well as other empirical studies. The quality of staff is emphasized in

[180] The fact that all dummies for subsectors (e.g. secondary education) are not significant is not supporting the fact that IEG reports on aggregate results of certain sub-sectors. This part can be explored in future studies.

Belassi and Tukel's (1996) framework as "factors related to project manager" and "project team members" stand out as separate categories that influence project success or failure. Factors related to project manager and project team members included elements, such as perception of his/her roles and responsibilities, competence, commitment, task team leader quality, technical background, team cohesion and troubleshooting skills. In addition to these qualifications of the project staff, Belassi and Tukel (1996) also discuss project manager's performance on the job that is influenced both by the staff qualifications and factors related to external environment and ultimately influence the project outcomes. The analysis of interview data from this study indicates that all the factors that are related to staff qualifications and performance on the job are important for project outcomes, thereby supporting the Belassi and Tukel's (1996) framework.

Middleton (1985) included "staffing and skills" and "staff continuity" as part of "management system factors" in his theoretical framework. The findings from this study strongly support the importance of these factors related to project staff. The number of staff members and their technical expertise needed in implementing projects should be adequate for project tasks. Moreover, the degree of key staff also matters for project outcomes as the interview data shows that frequent change of task team leaders and the local staff may lead to project failures. One of the efforts to improve the impact of World Bank staff is decentralization of the World Bank staff to borrowing countries in the 1990s. Recruiting local staff to support World Bank funded education programs have likely helped improving project delivery.

The importance of staff quality on project performance is also empirically supported by Denizer et al.'s (2013) study. Denizer et al. (2013) conducted task team

leader (at the World Bank) fixed effects and concluded that task team leader's quality is strongly associated with project outcomes. They found that task team leader fixed effects are of comparable importance to that of country fixed effects in accounting for the variation of project outcomes. Their study did not specifically point out the specific elements of task team leader quality that significantly influence project outcomes and the "task team leader" is limited to the World Bank staff. The findings from my study expand the findings from Denizer et al.'s (2013) study that not only the World Bank but also the local staff quality is important, and staff continuity, commitment, personality as well as expertise in the education projects are important factors for project outcomes. More specifically, lack of incentives for staff and management of partner donors can result in lack of aid effectiveness as Bermingham et al. (2009) pointed out. In fact, most development organizations reward staff for preparing projects and programs and there are a few rewards for more time consuming work, such as building effective partnerships with other agencies (Accra Agenda for Action, 2008). Future studies can explore more on the staff incentives and reward system within development agencies and project effectiveness. Moreover, future studies should also explore what aspects of the quality of task team leaders at the World Bank are specifically associated with better project outcomes, and the possible influence of task team leaders on project outcomes should be examined in a candid way.

The interview data implied that project design is also a crucial determinant for education project outcomes. A good design of an education project would be a project that fits the local context and needs, designed by an education expert and one that has a focused target. Project design is pointed out by Middleton's (1985) framework in the

"task factors" category, expressed as goals, agreements, complexity in terms of components and coordination, innovation and demand. Project design was also pointed out as a determinant for project outcomes by Ika et al. (2012; 2010). Ika et al.'s (2010) study of determinants for international development projects in Africa illustrated that project design combined with well monitored and evaluation tools are more likely to succeed. Another study by Ika et al. (2012) that analyzed World Bank projects found that among the factors that turned out to be significant for project success, project design, in addition to monitoring, were particularly significant components for project success.

Contrary to the conclusion from Ika et al.'s (2012) study, the interview data in my study reveals that no single factor is particularly more important than other components of the project. For example, project design was relevant to both countries of Maldives and Senegal, but due to other factors, the project implemented in Maldives received highly satisfactory ratings while the one in Senegal received a highly unsatisfactory rating. For the Maldives project, not only was the project relevant to the national priority, but also the government commitment, ownership and the staff continuity led to project's success (World Bank, 2002). On the other hand, a higher education project goals, implemented in Senegal, aligned with their national priorities, but due to unfavorable political economy factors combined with the lack of staff continuity and their weak implementation capacity, the project turned out to be a failure, receiving a highly unsatisfactory rating (World Bank, 2003b). These specific examples tell that there is not a single factor that matters the most for project outcomes. However, government willingness to carry out a project successfully seems to matter the most, since it means the willingness to control all the circumstances, which endanger the project. For example, if there were student riots on

campus that will clearly delay or ruin a project, the government can choose how to deal with this problem. If the government deals with these external factors that affect the project outcomes sufficiently with vigor, it is highly possible that a project can turn into a success despite all the difficulties. Thus, while there is no factor that stands out to heavily influence project outcomes, government commitment and willingness can still play a significant role even in circumstances that are difficult for project implementation.

New framework

The multiple factors that are defined from the interview data analysis in this study and the country, project level characteristics used in the quantitative study together illustrate a larger picture of the combination or relationships between these factors. The quantitative data did not show any evidence that simple characteristics of education projects such as project size, duration, loan types are associated with project ratings. On the contrary, the data suggested that government effectiveness is an important determinant for project outcomes. The interview data also supports this finding; however, the data suggests a clearer picture of the relationship between the determinants of project outcomes.

Through the interview data, I found that there are indirect factors and mediators of project outcomes. While operational management factors more directly influence project outcomes, environment factors, such as political instability, indirectly influence project outcomes through government personnel and relationship with the Bank. Although Middleton's framework illustrates that environment factors influence outcomes through project design, management and operational management factors, based on the evidence

from my study, I suggest a new framework that allows for environmental factors to directly influence project design, management and operational management factors simultaneously. Environmental factors only affect project outcomes through the mediators at the design, management and operational management level. For example, the political instability may lead to change of government, which leads to change in project implementation staff and when combined with a weak mechanism or weak government capacity to implement projects, this could lead to delays in the implementation of projects. A government with strong commitment to the project may be more likely to lead a smoother implementation of projects (easiness in delivery and procurement) that can result in receiving better ratings. Therefore, environment factors indirectly influence project outcomes.

Also, Middleton did not mention government ownership in the environment factor, however, based on evidence in my study, I emphasize the role of the government in the new framework. The significant finding of government effectiveness from the quantitative data indicates that the capacity of the government in delivering a public policy (in this case, education project) matters for project outcomes. The interview data also underlined the critical role of government for project success. The government needs to have strong political will, be committed to implement the project and put priority on education or on the project itself. In the new suggested framework, corruption and regional context is also emphasized in the category of environment factors.

I also took out "The nature of Bank support" from Middleton's framework from the category of management system factors and made a separate factor between management system and operational management factors. Since the qualitative data

indicated that the changed personnel of the government may lead to a change in the relationship between the government and the Bank, which can then lead to better or worse input management which ultimately affects the outcome, "The nature of Bank support" needs to be separated from the management system factors.

As Figure 7 illustrates, the direct factors for project outcomes are factors related to operational management, particularly input management such as delivery and procurement of project resources. Given the fact that the project performance ratings are evaluated based on the extent to project objectives were achieved, it is certain that the fewer issues related to procurement and delivery of activities, the more the number of planned project objectives will be achieved, which result in better evaluation ratings. On the other hand, if there are issues with input management such as delays of project input deliveries, less project objectives will be achieved, which then will result in lower project performance ratings. Jones (2007) explained that there could be a variety of reasons for delays of project input, including inadequate preparation, borrowers' difficulties with Bank procurement procedures, particular country problems of funding, social or political instability, and natural disasters. Therefore, lower ratings mean delays or issues involved with implementation of projects that can result from various factors at the management system, project design (preparation) or project environment level.

Overall, the quantitative and qualitative data in this study together shows that all categories of Figure 7 are important determinant for project outcomes. It goes without saying that a good combination of all of the factors in Figure 7 leads to project success. For example, a perfectly designed project with staff that has strong technical expertise to implement the project can still receive unsatisfactory ratings if the political context is not

favorable for project implementation. However, this study attempts to find factors that have a larger influence on project outcomes. This study shows evidence that the government ownership and regional context from the environment factor stand out as important determinants. The complexity of project objectives from the design factor, staff quality and continuity from the management system factors also stand out as crucial determinant for education project outcomes. Environment factors affect project outcomes through other factors at the design, management system and operational management level.

Scientific validity of project performance ratings

Although in some studies the researchers have argued that IEG outcomes are acceptable measures of program success, the validity of the IEG ratings emerged during the data collection and analysis in this study. The interview data analysis in this study shows both positive and negative views on the validity of project performance ratings. During the interviews with former and current education staff, many raised an issue regarding the measurement used for evaluating project outcomes. The ICR and IEG ratings were criticized to be too mechanical, not reflecting the context, incorrect indicators to measure the outcomes, and some personal reasons that may bias the evaluation[181]. As the interviewees suggested the shortcomings of IEG ratings, they

[181] The doubts on the validity of IEG ratings data does not necessarily lead to questions on the possibility for staff to manipulate indicators for project outcomes. First, as each project is implemented at least more than three years, the person who designed the project (and set up the project objectives, results framework, and indicators for project outcomes) is never held accountable for project results (Interview #3). Secondly, the ICR report is written not by the same person who were in charge of implementing the project, but by a third person who is never involved in the project implementation. This leads to little incentives to manipulate the indicators of success. Thirdly, the IEG evaluates that the ICR report is supported by a proof of evidence. This system makes it difficult for one to manipulate the indicators. Overall, the accountability within the World Bank is at the project level, not at the staff

emphasized that improved measure for evaluation should consider institutional building and sustainability of the project. Institutional building has long been recommended by the World Bank to borrowers. The OED report in 1978 recommends that the Bank should play a less prominent role in designing borrower's education system but rather the Bank's appropriate role is an institution builder and enhancing the borrower capacities to develop, manage, and evaluate their education and training systems (World Bank, 1978). Although the ICR report addresses institutional building and sustainability and the IEG data has ratings for sustainability, the interviewees argued that there needs to be a more formal measure to include these factors into final evaluation ratings. At the same time, there were also positive views on the ICR and IEG ratings. Interviewees have mentioned that it is useful, designed to be as objective as possible and definitely correlated to the reality; although they do not reflect the outcomes in reality perfectly.

Previous studies that I searched for this study all have viewed the ratings all from positive perspectives. Negative views on IEG ratings were not pointed out by any of the previous studies that used IEG ratings. Vawda et al. (2003) found a strong relationship between economic analysis (cost-benefit, cost-effectiveness analysis) and education project outcomes, thereby indirectly supporting that IEG ratings are valid measures for project success. Smets et al. (2012) have mentioned that the interviews of experienced IEG staff members supported the argument that there is no certain evidence to believe that IEG ratings are biased. Dollar and Svensson (2000) admitted that there is some subjective element in the assessment, however, since OED (now IEG) is independent from Bank's senior management (with a separate budget), there is no necessary bias in the results. However, the evidence of these arguments is derived from the IEG's internal

level. This should lead to little incentives for the staff to manipulate the indicators for results.

sources, except for Vawda et al.'s (2003) study that compared the project ratings to economic analysis. The validity of ratings that emerged from my study is also based on the evidence within the World Bank, yet it is from the Bank management staff, not from the IEG.

In sum, although the IEG ratings are credible to an extent, the findings in this study raise the question on the validity of the ratings data, which is not discussed in previous studies. More specifically, the findings suggest that the project performance ratings need to reflect the context of each project, carefully choose appropriate indicators to measure the project outcomes and incorporate sustainability and institutional building more systematically. Addressing the need to measure sustainability and institutional building as part of the outcome is particularly important for education projects where long-term goal is more important than other sector projects. This study also suggests that future studies that use IEG data need to test the scientific validity of the ratings. However, considering the context and nuance of each education project when evaluating will not be easy as these elements are difficult to be quantified. Therefore, as the results from quantitative data analysis in this study shows, more detailed data on each of the project needs to be collected to improve this study.

The fact that there is little match between quantitative and qualitative findings itself also raises doubt on the validity of evaluation ratings. For example, corruption was emerged during the qualitative data analysis. Despite the possibility that corruption during the implementation process can largely influence the project success, the IEG data does not carry any information related to corruption. This study, therefore, has limitations in explaining the influence of corruption on project outcomes. To improve a project

evaluation data, one should collect and analyze quantifiable data on corruption as well as data on other factors (e.g. project design) that stood out as important in qualitative data but not in quantitative data of this study.

CHAPTER VIII

CONCLUSION

This study aimed to examine the determinants of World Bank education project success using a mixed method design. For quantitative data analysis, I used the World Bank project performance ratings (ICR and IEG) data and other indicators at the country level and applied logit and ordinal logit models to conduct analyses. I found that country level characteristics, such as government effectiveness, regional dummies and education spending, were significantly associated with project ratings; however, I did not find any significant effect of project level characteristics on project success. For the qualitative data analysis, I interviewed 15 former and current education sector staff and found that project design related to project objectives, government ownership and commitment and quality of implementation staff were important for education project success. In addition, I found that the validity of project evaluation ratings emerged from the interview data that leads to a call for more scientific validation of project performance ratings.

The findings from both the quantitative and qualitative data in this study suggest that for education development projects, country characteristics, more specifically government leadership, commitment, ownership and effectiveness are significant factors for education project success. Government ownership and commitment is widely recognized as an important factor among the Bank staff, and also emphasized by the international community through the Paris Declaration on Aid Effectiveness, however, was not addressed in the previous studies. Since the education sector can easily be less of

a priority for governments of developing countries (with an exception of East Asian countries where education is highly valued), this study suggests that government leadership and commitment should be emphasized for the implementation of future educational development projects.

The finding that a country located in East Asia Pacific region has a significantly higher probability to receive satisfactory rating implies that some factors are outside the control of the Bank or the borrower government. For East Asian countries where education is highly valued, it may be very natural that education projects are rated better than education projects in other regions where education is less of a priority. However, my study shows that even though located in East Asia region, if the borrower has low government effectiveness, the probability of receiving a satisfactory rating will be lower than a borrower with high government effectiveness located in other region. Therefore this finding suggests that countries can improve the implementation of education projects if they have an effective government and the effectiveness of the government will overcome the factors not under their control, such as regional factors.

Regarding the project level determinants of project success, the quantitative data showed no evidence that any project level factor is critical determinant for project outcomes. Nevertheless, the interview data showed three factors perceived as important 1) staff continuity and skills 2) project design and 3) institutional capacity/government leadership. The finding that staff quality and project design are important factors support the previous studies on development projects (Khan, 1995; Ika, Diallo, & Thuillier, 2012; Denizer, Kaufmann, & Kraay, 2013; Ika, Diallo, & Thuillier, 2010). I also found that some additional factors such as relationship between the Bank and the government and

corruption stood out as significant factors for project outcomes. These factors were included in the previous theoretical model of World Bank education project performance suggested by Middleton (1985), therefore, the findings were expected. However, the interview data in this study described the detailed elements of each category that Middleton (1985) stated. Factors such as corruption were not mentioned in the theoretical framework, but frequently discussed by the interviewees in this study. This study also clarified the relationship between the elements and the categories. Middleton's (1985) framework included the Bank support within the category of management system factors; however, based on my data analysis, the Bank support needed to be separated out from the management system factor. Overall, the project level factors in the IEG ratings data did not show any significant effect on project outcomes, while the qualitative data showed that factors mentioned in previous literature (project design, quality of staff) were important as well as the factors that are specific to the World Bank projects (Bank support).

Lastly, the findings of this study raised a question on the validity of the IEG ratings. The interviewees were concerned about using the IEG ratings as a measure for project success. They had both positive and negative opinions on IEG ratings, but they generally agreed that the ratings do reflect the reality although not a perfect measure of project outcomes. Suggestions regarding ways improve the ratings also emerged from the data, such as including the measures of sustainability and institutional building. The findings from this study raise a new question on the credibility of project performance rating as a way to measure project success while previous studies showed little concerns for validity of IEG ratings.

This study is not without limitations. First, in terms of methodology, the quantitative part of this study has several limitations as the data is not detailed enough to conduct a sophisticated analysis. Using logistic regression analysis, the relationship between key interest variables and project ratings are correlational and not causal. This means that increased government effectiveness will not necessarily lead to better project outcomes, making cautions for policy recommendations. Detailed data on each of the project is needed for future studies. One way to do this is by coding the ICRs and seeking ways to quantify the qualitative information in the ICRs. Good quality data on each of the project will allow future researchers to conduct analysis with more advanced statistical methodology. Another limitation regarding the quantitative methodology in this study is that the results were sensitive to model specifications, making it difficult to make firm arguments on the government effectiveness as well as other country and project level variables. Future studies will need to provide more substantial evidence on the determinants of education project success. Lastly, as mentioned in the previous section, there may be potential omitted variable bias making the interest variable endogenous. This issue can be addressed by gaining data with more detailed information.

The qualitative part of this study may also have some limitations since the number of interviewees is not large. Due to rather small number of interviewees (15 interviewees) the findings from this study may be limited for its generalization in other contexts. There is also a subjective component of the interview data, as one would be heavily influenced by the projects that they were involved in. Although this may not be a major issue considering the total number of staff in education sector at the World Bank, future studies need to consider conducting a survey of former and current education sector staff.

Futures studies also need to examine how results change using different sequential mixed methods approach. Due to time constraints, in this study I first conducted quantitative analysis and then moved on to conduct qualitative analysis. However, if more detailed information on project characteristics is available for researchers, one should conduct qualitative part first and find what factors might influence project success. From the qualitative analysis one can find important variables in the quantitative data that may lead to better empirical design and obtain some significant findings in the quantitative analysis.

This study focuses education projects only, therefore it is not possible to compare whether education sector is more or less successful than projects in other sectors. There may be some project characteristics that are common across projects in all sectors rather than being uniquely relevant to education sector. Although discussing overall education sector success is beyond the scope of this study, future studies may need to broaden the sample size and examine how education projects are different from other sector projects.

The findings and implications of this study are based on donor (the World Bank Group) data and their perspective on education projects. It would be most ideal to reflect data and views from recipient/borrower country's side as well as that of beneficiaries. However, current research on international educational development is very limited to the data collected by developed countries/donors. The fact that the data is collected by donors implies that there may be a bias in determining factors that matter for project outcomes. For example, although the finding of this study show that borrower country's government effectiveness and commitment are significantly related to project outcomes, it is possible that the borrower country would perceive that Bank support may be the most significant

factor for carrying out projects successfully. Therefore, future studies need to address the perspectives from the borrower countries and beneficiaries of educational development projects. In this regard, the Bank is already conducting beneficiary assessment using qualitative data collection techniques and quantifying the findings (World Bank, 2015a).

Despite the limitations, the findings of this study call for a closer examination of the role of borrower country's government in project implementation. Rather than the amount or increase/decrease in aid, a more important question is how to deliver aid effectively and how to sustain the effect of development projects. To maximize a long lasting effect of education aid, the donors should focus on helping the governments to develop capacity in implementing education development projects successfully. In other words, the donors should provide more help on technical assistance rather than providing larger amounts of aid. The findings from this study support the need for more assistance from the donors in improving government effectiveness and motivating the governments to be more committed to education projects.

This study also contributes to the literature on aid effectiveness as it examines the aid effectiveness at the project level not at the country level. This study is also one of the few studies to focus specifically on the education sector with a mixed method research design that was not applied in other studies with similar topic. By using a mixed methods design, I attempted to capture both the significant determinants for project outcomes at the aggregate level as well as nuances, contexts and details of education project implementation. The results from both quantitative and qualitative data analysis imply that aid issue is particularly complicated in education sector and requires detailed data to conduct any research and make policy recommendations based on evidence from hard

data. The motivation, distribution, and effects of aid are complex and shifting, therefore, capturing this complexity requires detailed data and sophisticated methods that allow researchers to make causal and descriptive inferences (Tierney, et al., 2011). The interview data in this study also implies that every education project is implemented within its unique context. Consequently, Middleton (1985) also explains that generally there is no single project management approach that is effective, due to the differences in cultural and administrative contexts. Therefore, researchers need to pay careful attention to the nuances and details of each project when analyzing determinants for project success. Policy makers should consider the varying contexts in which projects are implemented, while the government effectiveness remains to be an objective in itself, the main purpose of public management sector.

APPENDIX

A. Definition and sources of each variable used in the analysis

Variable Name	Definition	Source
Outcome variable (IEG project rating)	The extent to which project's objectives have been achieved efficiently	World Bank IEG
Primary school gross enrollment	Ratio of total number of students who are enrolled in primary school, regardless of age, compared to the number of official primary school age	World Development Indicator
Education spending	Public expenditure on education as percentage of GDP	World Development Indicator
GDP growth	Annual GDP growth (in %)	World Development Indicator
Government effectiveness	Perceptions of the quality of public services, the quality of civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, the credibility of the government's commitment to such policies	Worldwide Governance Indicators
Region	Region where a project was implemented categorized by the World Bank	World Bank IEG
Project length	Length of time needed to implement a project (measured in years)	World Bank IEG
Project cost	Estimated project cost that includes non-Bank funding (in USD millions)	World Bank Projects and Operations Database
IDA (=1, IBRD=0)	Types of funding source within the Bank. IDA borrowers are usually least developed countries whereas IBRD borrowers are middle income countries	World Bank IEG
Development Policy Loans (=1, Specific Investment Loans=0)	Type of lending instrument. Investment loans have long-term focus (5-10 years), and Development Policy Loans have short-term focus (1-3 years).	World Bank IEG
Emergency Recovery Loan (=1, Others=0)	Supports economic and social recovery immediately after events such as war, civil disturbance, or natural disaster	World Bank IEG
Specific Investment Loan	Supports creation, rehabilitation, and maintenance of economic, social, and institutional infrastructure	World Bank IEG
Learning and Investment Loan	Loans of $5 millions or less financing	World Bank IEG

	small, experimental, risky and/or time-sensitive projects to pilot initiatives or to develop locally based models prior to large-scale intervention	
Technical Assistance Loan	The loan is to build institutional capacity in the borrower country.	World Bank IEG
Structural Adjustment Loan	Lending meant to catalyze policy reforms and that does not go to fund specific investments (Winters & Streitfeld, 2013)	World Bank IEG
Education sector (=1, Other sector=0)	Project that belong to Education Sector Board	World Bank Projects and Operations Database
Primary education (=1, other subsectors=0)	Any project in the data with more than 20 percent of the component from primary education sector	World Bank Projects and Operations Database
Secondary education (=1, other subsectors=0)	Any project in the data with more than 20 percent of the component from secondary education sector	World Bank Projects and Operations Database
Tertiary education (=1, other subsectors=0)	Any project in the data with more than 20 percent of the component from tertiary education sector	World Bank Projects and Operations Database
Vocational (=1, other subsectors=0)	Any project in the data with more than 20 percent of the component from vocational education sector	World Bank Projects and Operations Database
General education (=1, other subsectors=0)	Any project in the data with more than 20 percent of the component from general education sector	World Bank Projects and Operations Database
Repeater project (=1, Non-repeater projects=0)	Any project that is in the second or third phase of its initial project	World Bank Projects and Operations Database
Share of largest subsector	The largest share of a project assigned to a single subsector	World Bank Projects and Operations Database

B. Regression results with and without country weights applied

	Model_1	Model_1_W	Model_2	Model_2_W	Model_3	Model_3_W	Model_4	Model_4_W
	OLS	OLS	Logit	Logit	Ologit	Ologit	OLS_FE	OLS_FE
Dependent variable	Sat/Unsat	Sat/Unsat	Sat/Unsat	Sat/Unsat	0-5 rating	0-5 rating	Sat/Unsat	Sat/Unsat
Project level characteristics								
Project length	0.12*	-0.01	0.69	-0.03	0.75*	0.11	0.25**	0.19*
	(0.06)	(0.09)	(0.40)	(0.46)	(0.34)	(0.52)	(0.08)	(0.08)
Project cost (logged, in USD millions)	0.00	-0.00	0.01	-0.01	-0.11	-0.10	0.01	0.02
	(0.02)	(0.02)	(0.13)	(0.15)	(0.10)	(0.12)	(0.02)	(0.03)
Development policy loan	-0.12	0.01	-0.84	-0.04	-0.19	0.64	-0.07	0.07
	(0.11)	(0.14)	(0.82)	(0.90)	(0.57)	(0.76)	(0.13)	(0.17)
IDA project	-0.00	0.08	0.01	0.53	-0.17	0.03	-0.08	-0.03
	(0.09)	(0.10)	(0.60)	(0.67)	(0.37)	(0.47)	(0.16)	(0.13)
Education sector	-0.10	-0.11	-0.58	-0.56	-0.38	0.00	-0.09	-0.10
	(0.07)	(0.08)	(0.39)	(0.51)	(0.30)	(0.44)	(0.08)	(0.08)
Repeater project	0.06	0.13*	0.46	0.93*	-0.21	0.05	0.05	0.07
	(0.04)	(0.06)	(0.29)	(0.39)	(0.21)	(0.27)	(0.06)	(0.06)
Share of largest subsector in pct (logged)	-0.02	-0.02	-0.14	-0.28	-0.05	-0.34	-0.05	-0.06
	(0.06)	(0.08)	(0.39)	(0.50)	(0.25)	(0.35)	(0.07)	(0.08)
Emergency Recovery Loan	0.11	0.26	0.91	1.62	1.30	1.41*	0.10	0.11
	(0.13)	(0.14)	(0.92)	(0.99)	(0.70)	(0.70)	(0.17)	(0.20)
Learning and Innovation Loan	-0.21	-0.19	-1.08	-1.12	-0.55	-0.43	-0.22	-0.22
	(0.12)	(0.14)	(0.73)	(0.87)	(0.51)	(0.62)	(0.13)	(0.11)
Specific Investment Loan	-0.07	-0.05	-0.50	-0.20	0.10	0.35	-0.10	-0.07

	(0.08)	(0.09)	(0.54)	(0.60)	(0.28)	(0.39)	(0.08)	(0.08)
Technical Assitance Loan	-0.31	-0.29	-1.60	-1.81	-0.85	-0.79	-0.49*	-0.53**
	(0.17)	(0.20)	(0.88)	(1.04)	(0.64)	(0.82)	(0.20)	(0.18)
Structural Adjustment Loan	-0.08	-0.06	-0.30	-0.20	-1.10*	-1.24*	-0.08	-0.09
	(0.09)	(0.13)	(0.45)	(0.65)	(0.44)	(0.63)	(0.13)	(0.12)
Primary	0.11	0.10	0.73*	0.70	0.54*	0.35	0.10	0.11
	(0.06)	(0.07)	(0.34)	(0.41)	(0.26)	(0.36)	(0.07)	(0.06)
Secondary	0.01	0.01	0.01	0.01	0.43	0.14	0.03	0.03
	(0.07)	(0.08)	(0.45)	(0.55)	(0.32)	(0.41)	(0.08)	(0.09)
Tertiary	-0.04	-0.04	-0.25	-0.33	0.43	0.08	-0.08	-0.18*
	(0.08)	(0.09)	(0.47)	(0.53)	(0.40)	(0.44)	(0.08)	(0.09)
Vocational	0.01	0.05	0.04	0.17	0.45	0.35	0.04	0.12
	(0.11)	(0.11)	(0.66)	(0.66)	(0.49)	(0.55)	(0.10)	(0.11)
General education	-0.05	-0.15*	-0.25	-0.89*	0.42	-0.25	-0.02	-0.08
	(0.07)	(0.07)	(0.38)	(0.45)	(0.35)	(0.43)	(0.07)	(0.08)
Country level characteristics								
Primary school gross enrollment (logged)	-0.32*	-0.29	-1.68*	-1.70	-1.23*	-1.37*	-0.21	-0.12
	(0.13)	(0.15)	(0.67)	(0.91)	(0.52)	(0.68)	(0.38)	(0.48)
GDP growth (annual %)	0.02	0.01	0.11*	0.09	0.07	0.04	0.01	0.00
	(0.01)	(0.01)	(0.05)	(0.06)	(0.04)	(0.05)	(0.01)	(0.01)
Education spending as % of GDP	-0.04*	-0.04*	-0.24	-0.27*	-0.12	-0.06	-0.04	0.01
	(0.02)	(0.02)	(0.12)	(0.12)	(0.08)	(0.09)	(0.05)	(0.06)
Government effectiveness	0.14*	0.13	0.93*	1.02*	0.54	0.47	0.08	0.13
	(0.07)	(0.07)	(0.46)	(0.47)	(0.29)	(0.36)	(0.25)	(0.27)
Regional dummies								

Africa	-0.02	-0.06	-0.03	-0.24	-0.02	-0.20		
	(0.11)	(0.11)	(0.59)	(0.63)	(0.41)	(0.50)		
East Asia Pacific	0.20	0.17	1.57**	2.06*	1.07*	1.11*		
	(0.10)	(0.09)	(0.55)	(0.82)	(0.49)	(0.56)		
South Asia	0.15	0.07	1.03	0.52	0.56	0.24		
	(0.11)	(0.10)	(0.64)	(0.71)	(0.39)	(0.47)		
Latin America and Caribbean	0.14	0.12	0.85	0.61	0.67	0.45		
	(0.13)	(0.12)	(0.67)	(0.66)	(0.53)	(0.59)		
Middle East and North Africa	0.09	-0.03	0.57	-0.05	0.20	-0.33		
	(0.15)	(0.16)	(0.81)	(0.83)	(0.56)	(0.61)		
Constant	2.55***	2.16**	10.99**	10.10*			2.44	1.64
	(0.69)	(0.77)	(3.87)	(4.63)			(1.72)	(2.27)
cut1								
Constant					-12.71***	-13.50***		
					(3.05)	(3.72)		
cut2								
Constant					-10.27***	-10.52**		
					(2.99)	(3.72)		
cut3								
Constant					-8.84**	-8.82*		
					(2.98)	(3.72)		
cut4								
Constant					-7.24*	-7.15		
					(2.98)	(3.70)		
cut5								
Constant					-3.60	-2.82		
					(2.98)	(3.74)		
Year dummies	Y	Y	Y	Y	Y	Y	Y	Y

Country fixed effects	N		N		N		N		N		N		Y		Y	
R sqared within														0.2		0.25
R sqared between														0.03		0.0014
R squared overall		0.19		0.25		0.18		0.25		0.08		0.1		0.11		0.07
Observations	404		404		404		404		404		404		404		404	

Note: Cluster standard errors in parentheses * p<0.05 ** p<0.01 *** p<0.001

Note: Pseudo R squared for logit and ordinal logit models

Note: "Model_#_W" indicates models with country weights (1/number of projects within a country) applied. Data used in the analysis is not imputed for missing data.

C. Informed consent

What makes good projects? Success factors of the World Bank education development projects

Consent for Bank staff's Participation in Research

April/May, 2014

Dear _____

Bommi Lee is a doctoral student from the department of Leadership, Policy and Organizations at Vanderbilt University. This letter asks you to work with her in a research study called "What makes good projects? Success factors of the World Bank education development projects." You are being asked to participate because you either have worked for or are currently working for the World Bank education sector. The purpose of this research is to investigate the factors that influence project outcomes.

What will I be asked to do?

1) Allow the researcher to interview you for about 30 minutes to an hour.
2) Allow the researcher to audio-record and take notes during the interview.

What will happen during the interview?

The researcher will interview you at a time that is convenient for you. The interview will last for about 30 minutes to an hour. The researcher will take notes and audio-record the interview. You will be asked about your perception on what factors are important to project outcomes and your perception of whether the project outcomes assessed in the Implementation Completion Report reflects your perceived success of the project.

Do I have to be in this study?

No, you do not have to take part in this research. Taking part in this research is completely voluntary. If you say "yes" now, you can change your mind later. You can drop out at any time by contacting Ms. Bommi Lee at 615-429-6153. You may also let her know that you would like to quit participating via e-mail at bommi.lee@vanderbilt.edu. Also, you may discontinue the interview at any time if you feel uncomfortable. Failing to complete the interview will not negatively affect you in any way.

Who will see the notes and listen to the recorded audio file?

All efforts, within reason, will be made to keep your information confidential. That means, I will not let other people see or hear the interview unless you say it is okay. Total confidentiality cannot be guaranteed. Only the researcher will have access to all the audio-recordings and notes. I will share some aggregate results from all of the interviews with other educators and researchers. I may also use part of the notes or audio recordings in the writings. I will make sure all interviewees are kept anonymous and use nicknames that are not identifiable of each interviewee. I will keep the written notes in a locked cabinet. The digital file of audio recordings will be kept in a locked file. Research information may also be kept on a computer protected with a password and encrypted files. All digital audio files will be discarded as soon as I transcribe them. The transcripts of the interviews will also be kept in a locked file.

Will others know that I am taking part in this research?
Your staff will not know that you are taking part in this research unless you tell them. To make it harder to identify you, I will use a fake name when writing about you in the dissertation or presentations. Your information may be shared with Vanderbilt, such as the Vanderbilt University Institutional Review Board, Federal Government Office for Human Research Protections, to check how I did the research, or if I am required to share the results by law.

What are the risks and inconveniences of taking part in this research?
I do not believe this study presents any serious risks. There may be a minor risk that someone might be able to identify you in the dissertation when presented to other researchers. However, as mentioned above, I will try to present the aggregate findings from the interviews rather than presenting findings from a specific interviewee.

What are the benefits of taking part in this research?
There is no anticipated benefit to you by taking part in this research. To show you the researcher's appreciation for participating in this research, the researcher will give you a small renumeration in the form of a gift card worth $20 (Starbucks).

If I want to participate, what do I have to do?
Please sign this letter and check the boxes that show your choices. Return it to Ms. Bommi Lee. Please keep a copy for your records.

What if I have questions?
If you have questions about the research, please contact **Bommi Lee** at **615-429-6153** or by e-mail bommi.lee@vanderbilt.edu or my Faculty Advisor**, Stephen Heyneman** at **615-322-1169**. If you have questions about giving consent to participate in the research, your rights as a participant, or want to offer input please call the Vanderbilt Institutional Review Board at (615) 322-2918 or toll free at (866) 224-8273.

Thank you very much!

Sincerely,

Bommi Lee
Doctoral Student
Dept. of Leadership, Policy and Organizations.
Peabody College, Vanderbilt University

D. Interview guide

Interview guide

Summary memo
Interviewee (Title and name): _____
Interview date/time and venue:_____
Main topics discussed: _____
Other topics discussed: _____
Post interview comments: _____

Introduction

My name is Bommi Lee, a Ph.D. student studying international education policy at Vanderbilt University. To facilitate my note-taking, I would like to audio record our conversations today. For your information, only researchers on the project will be privy to the audio file, which will be eventually destroyed after they are transcribed. In addition, you must read a consent form and sign if you agree. Essentially, this document states that: (1) all information will be held confidential, and (2) your participation is voluntary and you may stop any time. Thank you for your agreeing to participate.

I have planned this interview to last no longer than one hour. During this time, I have several questions that I would like to cover. If time begins to run short, it may be necessary to interrupt you in order to push ahead and complete this line of questioning.

You have been selected for being an interviewee because you have been identified as someone who has a great deal to share about the planning, implementation, and evaluation of education projects at the World Bank, with particular interest in the evaluation process and factors that might influence the project outcomes. My study does not aim to evaluate your techniques or experiences. Rather, I am trying to learn more about how education projects are assessed by the Bank staff, and hopefully discover some important factors that are unseen from the available IEG data that help a project to receive a more satisfactory outcome.

A. Interviewee background/work experience

Mr. or Ms. ____, I'd like to begin by asking you to explain briefly about your educational background and how long you have been working for the World Bank?
* Probing questions
– What were your tasks and responsibilities in the education sector? / What kind of work (projects) were you involved in?
– Could you explain in detail about one specific project.
– Do you have experience in writing a project completion report? (describe in detail about the evaluation procedure)

B. Perception on project success/failure

How did the project end?
When you rate a project as "successful" or "unsuccessful" in the completion report, would you say they reflect the actual project outcome well? or do you have any other reasons, (e.g. organizational pressure) to rate a project as "successful"? – How do you view the project evaluation system at the Bank?

C. Perceived factors that significantly influence project outcomes

If the project you described was “successful” or “unsuccessful,” what do you think is the reason for it?

When you write Project Completion Reports and give overall ratings of projects, what is the most influential factor that determines your ratings?

REFERENCES

Accra Agenda for Action (AAA). (2008). 3rd High level forum on aid effectiveness, September 2–4, 2009. Accra, Ghana.

Akhand, H. A., & Gupta, K. L. (2002). *Foreign aid in the twenty-first century.* Springer US.

Allison, P. (2009). *Fixed effects regression models* (Vol. 160). Sage Publications.

Allison, P. (2013). *What's the best R squared for logistic regression? Statistical horizons*. Retrieved 2015, from http://www.statisticalhorizons.com/r2logistic

Andersen, T. B., Hansen, H., & Markussen, T. (2006). US politics and World Bank IDA lending. *Journal of Development Studies, 42* (5), 772–794.

Baum, C. F. (2006). *An introduction to modern econometrics using Stata.* College Station, Texas: A Stata Press Publication. StataCorp LP.

Baum, W. C. (1982). *The project cycle.* Washington, DC: The World Bank.

Belassi, W., & Tukel, O. I. (1996). A new framework for determining critical success/failure factors in projects. *International Journal of Project Management, 14* (3), 141–151.

Benveniste, L. (1999). The politics of student testing: A comparative analysis of national assessment systems in southern cone countries. *Ph.D thesis*. Stanford University School of Education.

Berg, E. (2000). Aid and failed reform: The case of public sector management. In F. Tarp, Foreign *aid and development: Lessons learned and directions for the future.* London, UK: Routledge.

Bermingham, D., Christensen, O. R., & Mahn, T. C. (2009). Aid effectiveness in education: Why it matters. *Prospects, 39*, 129–145.

Birchler, K., & Michaelowa, K. (2013). *Making aid work for education in developing countries:* An *analysis of aid effectiveness for primary education coverage and quality.* UNU-WIDER.

Birdsall, N., & Savedoff, W. (2010). *Cash on delivery: a new approach to foreign aid.* Washington, DC: Center for Global Development.

Bollag, B. (2004). Improving *tertiary education in Sub-Saharan Africa: Things that work.* Africa Region Human Development Working Paper. Washington, DC: World Bank.

Boone, P. (1994). The *impact of foreign aid on savings and growth.* Centre for Economic Performance. London School of Economics.

Bourguignon, F., & Sundberg, M. (2007). Aid effectiveness: Opening the black box. *American Economic Review, 97*, 316–21.

Brautigam, D., & Knack, S. (2004). Foreign aid, institutions, and governance in Sub-Saharan Africa. *Economic Development and Cultural Change, 52*.

Bruckner, M. (2009, September). *Getting the effect of foreign aid on economic growth right: An issue of* addressing *endogeneity bias.* Retrieved September 8, 2015, from Universitat Mannheim: https://www2.vwl.uni-mannheim.de/fileadmin/user_upload/pigorsch/pdf/brueckner_aid_and_growth.pdf

Burnett, N., & Felsman, C. (2012). *Post-2015 Education MDGs.* Washington: Results for Development Institute.

Burnside, C., & Dollar, D. (1997). Aid spurs growth-in a sound policy environment. *Finance and Development, 34*, 4–7.

Burnside, C., & Dollar, D. (2000). Aid, policies, and growth. *American Economic Review, 90* (4), 847–868.

Burnside, C., & Dollar, D. (2004). Aid, policies, and growth: Reply. *American Economic Review, 94*, 781–4.

Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using Stata.*

Carvalho, S., & White, H. (1997). *Combining the quantitative and qualitative approaches to poverty measurement and analysis.* Washington DC: World Bank.

Cassen, R. (1994). *Does aid work? Report to an intergovernmental task force.* OUP Catalogue.

Chabbott, C. (2003). *Constructing education for development: International organizations and education for all.* Routledge.

Charmaz, K. (2006). *Constructing grounded theory.* London: Sage.

Chauvet, L., Collier, P., & Duponchel, M. (2010). *What explains aid project success in post-conflict situations?* World Bank.

Chauvet, L., Collier, P., & Fuster, A. (2006). *Supervision and project performance: A principal agent approach.* Retrieved from http://basepub.dauphine.fr/xmlui/handle/123456789/14817

Christensen, Z., Homer, D., & Nielson, D. L. (2011). Dodging adverse selection: How donor type and governance condition aid's effects on school enrollment. *World Development, 39* (11), 2044–2053.

Clemens, M. A., Radelet, S., & Bhavnani, R. R. (2004). *Counting chickens when they hatch: The short term effect of aid on growth.* Center for Global Development Working Paper 44.

Clemens, M., & Radelet, S. (2003). *The millennium challenge account: How much is too much, how long is long enough?* Washington: Center for Global Development.

Collier, P., & Dehn, J. (2001). *Aid, shocks, and growth.* Washington, DC: World Bank.

Collier, P., & Dollar, D. (2004). Development effectiveness: What have we learnt? *The Economic Journal, 114* (496), F244-F271.

Corbin, J., & Strauss, A. (2008). *Basic of qualitative research* (3rd Edition ed.). Los Angeles: Sage.

Cordella, T., & Dell'Ariccia, G. (2007). Budget support versus project aid: A theoretical appraisal. *The Economic Journal, 117*, 1260–1279.

Creswell, J. (2003). *Research design. Qualitative, quantitative, and mixed methods approaches* (2nd Edition ed.). Thousand Oaks, London, New Delhi: Sage Publications.

Dalgaard, C. J., & Hansen, H. (2001). On aid, growth and good policies. *Journal of Development Studies, 37* (6), 17–41.

de Siqueira, A. C. (2012). The 2020 World Bank education strategy. In S. Klees, J. Samoff, & N. P. Stromquist, *The World Bank and education: Critiques and alternatives* (pp. 69–81). Rotterdam, Boston, and Taipei: Sense Publishers.

Deininger, K., Squire, L., & Basu, S. (1998). Does economic analysis improve the quality of foreign assistance? *The World Bank Economic Review, 12* (3), 385–418.

Denizer, C., Kaufmann, D., & Kraay, A. (2013). Good countries or good projects? Macro and micro correlates of World Bank project performance. *Journal of Development Economics, 105*, 288–302.

Diallo, A., & Thuillier, D. (2005). The success of international development projects, trust and communication: An African perspective. *International Journal of Project Management, 23* (3), 237–252.

Dollar, D., & Levin, V. (2005). *Sowing and reaping: Institutional quality and project outcomes in developing countries.* World Bank.

Dollar, D., & Svensson, J. (2000). What explains the success or failure of structural adjustment programmes? *The Economic Journal, 110* (466), 894–917.

Doucouliagos, H., & Paldam, M. (2009). The aid effectiveness literature: The sad results of 40 years of research. *Journal of Economic Surveys, 23* (3), 433–461.

Dreher, A., Klasen, S., Vreeland, J. R., & Werker, E. (2010). *The costs of favoritism: Is politically driven aid less effective?* Insitute for the Study of Labor (IZA).

Dreher, A., Nunnenkamp, P., & Thiele, R. (2008). Does aid for education educate children? Evidence from panel data. *The World Bank Economic Review, 22* (2), 291–314.

Dreher, A., Sturm, J. E., & Vreeland, J. R. (2009). Development aid and international politics: Does membership on the UN Security Council influence World Bank decisions? *Journal of Development Economics, 88* (1), 1–18.

Easterly, W. (2006). *The white man's burden: Why the west's efforts to aid the rest have done so much ill and so little good.* New York: Penguin Press.

Easterly, W., Levine, R., & Roodman, D. (2004). Aid, policies, and growth: Comment to Burnside and Dollar, 2000. *American Economic Review, 94*, 774–780.

Erlandson, D. A., Harris, E. L., Skipper, B. L., & Allen, S. D. (1993). *Doing natrualistic research. A guide to methods.* Newbury Park, CA: Sage.

Fleck, C., & Kilby, R. (2005). World Bank independence: A model and statistical analysis of U.S. influence. *Review of Development Economics, 10* (2), 224–240.

Fredriksen, B. (1990). *Increasing foreign aid for primary education: the challenge for donors.* . Education and Employment Division, Population and Human Resources Department, the World Bank.

Fredriksen, B. (2011a). Aid dependency risks in the education sector: A review of issues. *Financing education: Redesigning national strategies and the global aid architecture* (pp. 185–206). Paris: International Institute of Education Planning.

Fredriksen, B. (2011b). Using aid strategically to promote education development: The need for a more effective global education aid architeture. *Optimizing talent:*

Closing educational and social mobility gaps worldwide. Salzburg, Austria: Salzburg Global Seminar.

Fredriksen, B., & Kagia, R. (2013). Attaining the 2050 Vision for Africa Breaking the Human Capital Barrier. *Global Journal of Emerging Market Economies*, *5*(3), 269-328.

Frey, B., & Schneider, F. (1986). Competing models of international lending activity. *Journal of Development Economics, 20*, 225–45.

Gavin, M., & Rodrik, D. (1995). The World Bank in historical perspective. *The American Economic Review*, 329–334.

Gilbert, C. L., & Vines, D. (2000). The World Bank: An overview of some major issues. *The World Bank: Structure and policies, 3*, 10.

Guillaumont, P., & Chauvet, L. (2001). Aid and performance: A reassessment. *Journal of Developmental Studies, 37* (6), 66–92.

Guillaumont, P., & Laajaj, R. (2006). *When instability increases the effectiveness of aid projects.* World Bank.

Hansen, H., & Tarp, F. (2000). Aid effectiveness disputed. *Journal of International Development, 12* (3), 375–98.

Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature, 46* (3), 607–68.

Hassan, M. (2012). External environment as a critical success factor in World Bank projects. *Thesis* . University of Colorado at Boulder.

Heller, P. S., & Gupta, S. (2002). *Challenges in expanding development assistance.* Washington DC: International Monetary Fund.

Heyneman, S. P. (1985). Diversifying secondary school curricula in developing countries: An implementation history and some policy options. *International Journal of Educational Development, 5* (4), 283–288.

Heyneman, S. P. (1995). Economics of education: Disappointment and potential. *Prospects, 25*, 559–583.

Heyneman, S. P. (1999). Development aid in education: A personal view. *International Journal of Educational Development, 19*, 183–190.

Heyneman, S. P. (2001). The World Bank and the analysis of human capital: A contribution or distortion? Washington DC: Annual Meeting of the Comparative and International Education Society.

Heyneman, S. P. (2003). The history and problems in the making of education policy at the World Bank 1960–2000. *International Journal of Educational Development, 23*(3), 315–337.

Heyneman, S. P., & Lee, B. (2013). *International organizations and the future of education assistance.* United Nations University-World Institute for Development Economics Research.

Ika, L. A., Diallo, A., & Thuillier, D. (2010). Project management in the international development industry: The project coordinator's perspective. *International Journal of Managing Projects in Business, 3* (1), 61–93.

Ika, L. A., Diallo, A., & Thuillier, D. (2012). Critical success factors for World Bank projects: An empirical investigation. *International Journal of Project Management, 30*(2012), 105–116.

Isham, J., Kaufmann, D., & Pritchett, L. (1997). Civil liberties, democracy, and the performance of government projects. *World Bank Economic Review, 11*(2), 219–242.

Isham, J., & Kaufmann, D. (1999). The forgotten rationale for policy reform: The productivity of investment projects. *Quarterly Journal of Economics, 114*(1), 149–184.

Jain, P. S. (1997). Program success and management of integrated primary education in developing countries. *World Development, 25* (3), 349–358.

Jelovac, I., & Vandeninden, F. (2008). *How should donors give foreign aid? Project aid versus budget support.* Working Papers 2008/04, Center for Public Economics and Population Economics (CREPP).

Jones, P. W. (1992). *World Bank financing of education: Lending, learning and development* (1st Edition ed.). Routledge.

Jones, P. W. (2007). *World Bank financing of education: Lending, learning and development* (2nd Edition ed.). Routledge.

Kaja, A., & Werker, E. (2010). Corporate governance at the World Bank and the dilemma of global governance. *The World Bank Economic Review, 24* (2), 171–198.

Kanbur, R. (2000). Aid, conditionality and debt in Africa. In F. Tarp, *Foreign aid and development: lessons learnt and directions for the future.* London and New York: Routledge.

Kaufmann, D., & Wang, Y. (1995). Macroeconomic policies and project performance in the social sectors: A model of human capital production and evidence from LDCs. *World Development, 23* (5), 751–765.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2006). *Governance matters V: Aggregate and individual governance indicators for 1996–2005.* Washington DC: World Bank.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). *The worldwide governance indicators: Methodology and analytical issues.* The World Bank.

Kennedy, K. J., & Lee, J. (2007). *The changing role of schools in Asian societies: Schools for the knowledge society.* Routledge.

Khan, J. (1995). An educational project in a developing country. *International Journal of Educational Management, 9*(1), 42–48.

Khang, D. B., & Moe, T. L. (2008). Success criteria and factors for international development projects: A life-cycle-based framework. *Project Management Journal, 39*(1), 72–84.

Kilby, C. (1994). *Risk management: An econometric investigation of project-level factors.* Backgroud paper for the World Bank's Operations Evaluation Department Annual Review of Evaluation Results.

Kilby, C. (2000). Supervision and performance: the case of World Bank projects. *Journal of Development Economics, 62* (2000), 233–259.

Kilby, C. (2001). World Bank borrower relations and project supervision. *Canadian Journal of Development Studies, 22* (1), 191–218.

Kilby, C. (2012). *Assessing the contribution of donor agencies to aid effectiveness: The impact of World Bank preparation on project outcomes.* No.20, Villanova School of Business, Department of Economics and Statistics.

Kilby, C. (2013). The political economy of project preparation: An empirical analysis of World Bank projects. *Journal of Development Economics, 105*, 211–225.

Killick, T. (1995). Conditionality and the adjustment-development connection. *Pakistan Journal of Applied Economics, 11*, 17–36.

Killick, T. (2003). Macro-level evaluations and the choice of aid modalities. *Paper prepared for OED conference on Evaluating Development Effectiveness.*

King, E. (2013). Education at the World Bank Group. Presentation to 1818 Society.

Koeberle, S., Bedoya, H., Silarszky, P., & Verheyen, G. (2005). *Conditionality revisited: Concepts, experiences, and lessons.* Washington, DC: World Bank.

Kopp, H. E. (2003). Promoting professional and personal trust in OED. In P. G. Grasso, S. S. Wasty, & R. V. Weaving, *World Bank Operations Evaluation Department: The first 30 years* (pp. 55–60). The World Bank Publications.

Laruch, L., & Lockheed, M. (1992). *World Bank lending for educational testing. Background paper 92/62R.* Population and Human Resources Development Department. Washington, DC: World Bank.

Lee, K. W. (2010). Borrowing from the World Bank for education: lessons from Korea and Mexico. *Journal of International Cooperation in Education*,*13*(2), 49-91.

Lensink, R., & White, H. (2001). Are there negative returns to aid? *Journal of Development Studies, 37*, 42–65.

Leung, F. (2001). In search of an East Asian identity in mathematics education . *Educational Studies in Mathematics, 47* (1), 35–41.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.

London, N. A. (1993). When education projects in developing countries fail: A case study. *International Journal of Educational Development, 13* (3), 265–275.

Lyman, R. (1998). Got a cause and a computer? You can fight city hall. *The New York Times*, p. A12.

Marshall, K. (2008). *The World Bank: from reconstruction to development to equity.* London and New York: Routledge.

Marshall, M. G., & Cole, B. R. (2011). *Global report 2011: Conflict, governance, and state fragility.* George Mason University: Center for Systemic Peace and Center for Global Policy.

Michaelowa, K., & Weber, A. (2006). Aid effectiveness in the education sector: A dynamic panel analysis. *Frontiers of Economics and Globalization, 1*, 357–385.

Michaelowa, K., & Weber, A. (2008). *Aid effectiveness in primary, secondary, and tertiary education. Background paper for the Education for All Global Monitoring Report 2008.* New York: UNESCO.

Middleton, J. (1985). *Management in World Bank education projects: Analysis of experience.* World Bank.

Moloney, K. (2009). Public administration and governance: a sector-level analysis of World Bank aid. *International Review of Administrative Sciences, 75*(4), 609–627.

Moock, R., & Harbison, R. W. (1988). *Education in Sub-Saharan Africa: policies for adjustment revitalization and expansion.* Washington, DC: World Bank.

Mosley, P., Harrigan, J., & Toye, J. (1995). *Aid and power: The World Bank policy based lending* (2nd Edition ed.). London and New York: Routledge.

Mundy, K. (2002). Retrospect and prospect: education in a reforming World Bank. *International Journal of Educational Development, 22*(5), 483–508.

Mundy, K. (2006). Education for All and the new development compact. *Review of Education, 52*, 23–48.

Mundy, K., & Verger, A. (2015). The World Bank and its evolving role as an education policy governor. In K. Martens, & A. Kaasch, *Actors and Agency in Global Social Governance.* Oxford: Oxford University Press.

Murphy, P. (1996). *National assessments: Testing the system. EDI Learning Resources Series.* Washington, DC: Distribution Unit, Office of the Publisher, The World Bank.

Nandy, K. (2012, April 9). Understanding and quantifying effect sizes. UCLA.

Nielsen, H. D. (2006). *From schooling access to learning outcomes, an unfinished agenda: an evaluation of World Bank support to primary education.* World Bank Publications.

Nissanke, M. (2010). Reconstructing the aid effectiveness debate. In G. Mavrotas, *Foreign aid for development: Issues, challenges, and the new agenda* (pp. 63–93). Oxford University Press.

OECD. (2005, March). *The Paris Declaration on Aid Effectiveness and the Accra Agenda for Action.* Retrieved 2015, from OECD: http://www.oecd.org/dac/effectiveness/34428351.pdf

OECD. (2015). *OECD.Stat*. Retrieved October 29, 2013, from OECD statistics: http://stats.oecd.org/

OECD. (n.d.). *The Paris Declaration on Aid Effectiveness: Five principles for smart aid.* Retrieved 2015, from OECD: http://www.oecd.org/dac/effectiveness/45827300.pdf

Overseas Development Institute. (2009). *A development charter for the G-20. Background paper.* Retrieved from Overseas Development Institute: http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/4144.pdf

Petrosino, A., Morgan, C., Fronius, T. A., Tanner-Smith, E. E., & Boruch, R. F. (2012). Interventions in developing nations for improving primary and secondary school enrollment of children:A systematic review. *Cochrane Database of Systematic Reviews, 19*.

Psacharopoulos, G. (1980). *Higher education in developing countries: A cost-benefit analysis.* Washington, DC: World Bank.

Psacharopoulos, G. (1986). *Financing education in developing countries: An exploration of policy options.* Washington, DC: The World Bank.

Psacharopoulos, G. (1990). *Why educational policies can fail: an overview of selected African experiences.* Washington, DC: World Bank.

Rajan, R. G., & Subramanian, A. (2005). *Aid and growth: What does the cross-country evidence really show?* International Monetary Fund. Washington, DC: IMF.

Ranis, G. (2010). Towards the enhanced effectiveness in foreign aid. In G. Mavrotas, *Foreign aid for development: Issues, challenges, and the new agenda WIDER Studies in Development Economics* (pp. 57–62). Oxford: Oxford University Press.

Research Gate. (n.d.). *Logistic regression on SPSS*. Retrieved 2015, from Research Gate: http://www.researchgate.net/publictopics.PublicPostFileLoader.html?id=5210561ccf57d7a42f3e3baa&key=e0b495210561cbba06

Roodman, D. (2007). *Macro aid effectiveness research: A guide for the perplexed.* Center for Global Development.

Rubin, H. J., & Rubin, I. S. (2012). *Qualitative interviewing: The art of hearing data* (3rd Edition ed.). Los Angeles, London, New Delhi, Singapore, Washington DC: Sage.

Samoff, J. (2012). More of the same will not do. Learning without learning in the World Bank's 2020 Education Strategy. In S. Klees, J. Samoff, & N. P. Stromquist, *The World Bank and education: critiques and alternatives* (pp. 109–124). Rotterdam, Boston, and Taipei: Sense Publishers.

Samoff, J., & Carrol, B. (2003). From manpower planning to the knowledge era: World Bank policies on higher education in Africa. *UNESCO Forum on Higher Education, Research and Knowledge.* Division of Higher Education, UNESCO.

Schutt, R. K. (2009). *Investigating the social world: the process and practice of research* (6th Edition ed.). University of Massachusetts, Boston: Pine Forge Press.

Smets, L., Knack, S., & Molenaers, N. (2012). *Political ideology, quality at entry and the success of economic reform programs.* The World Bank Development Research Group.

Steer, L., & Smith, K. (2015). *Financing Education: Opportunities for Global Action*. Center for Universal Education at Brookings.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research. Grounded theory procedures and techniques.* Newbury Park, CA: Sage.

Svensson, J. (1999). Aid, growth and democracy. *Economics and Politics, 11*, 275–297.

Svensson, J. (2003). Why conditional aid doesn't work and what can be done about it? *Journal of Development Economics, 70*, 381–402.

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 19* (3), 227.

Tarp, F. (2010). Aid, growth and development. In G. Mavrotas, *Foreign aid for development: Issues, challenges, and the new agenda. WIDER Studies in Development Economics* (pp. 20–53). Oxford: Oxford University Press.

Task Force on Higher Education. (2000). *Higher education in developing countries: Peril and promise (No.440).* World Bank.

Teddlie, C., & Tashakkori, A. (2009). *The foundations of mixed methods research: Integrating quantitative and qualitative techniques in the social and behavioral sciences.* Thousand Oaks, CA: Sage.

Tierney, M. J., Nielson, D., Hawkins, D. G., Roberts, J. T., Findley, M. G., Powers, R. M., et al. (2011). More dollars than sense: Refining our knowledge of development finance using AidData. *World Development Journal*, *39*(11), 1891–1906.

Transparency International. (2015). *Corruption Perceptions Index*. Retrieved 2015, from Transparency International: http://www.transparency.org/research/cpi/overview

UNESCO. (2012). *EFA Global Monitoring Report 2012: Youth and skills. Putting education to work.* Paris: UNESCO.

UNESCO. (2013a). *Education for All Global Monitoring Report 2013/4 Teaching and Learning: Achieving Quality for All.* UNESCO.

UNESCO. (2013b). *Trends in aid to education: Lessons for post-2015.* UNESCO Education for All Global Monitoring Report. UNESCO.

United Nations. (2013). *A new global partnership: eradicate poverty and transform economies through sustainable development. The report of the High-Level Panel of Eminent Persons on the post-2015 development agenda.* United Nations.

Vawda, A. Y., Moock, P., Gittinger, J., & Patrinos, H. A. (2003). Economic analysis of World Bank education projects and project outcomes. *International Journal of Educational Development, 23* (6), 645–660.

Verspoor, A. (2001). *A Chance to Learn: Knowledge and Finance for Education in Sub-Saharan Africa. Africa Region Human Development Series*. World Bank Publications, PO Box 960, Herndon, VA 20172-0960.

Wang, C., & Bergquist, M. (2003). Basic education development in China: from finance reform to World Bank projects. *The International Journal of Educational Management, 17*(7), 303–311.

White, H. (2008). Of probits and participation: The use of mixed methods in quantitative impact evaluation. *IDS bulletin, 39*(1), p. 98.

Williams, R. (2015, February 21). *Ordered logit models-overview.* Retrieved March 20, 2015, from University of Notre Dame Sociology 73994 Categorical Data Analysis: https://www3.nd.edu/~rwilliam/stats3/L11.pdf

World Bank. (1974). *Education sector policy paper.*

World Bank (1978) Review of Bank Operations in the Education Sector (the "Romain Report"), Washington, DC: Operations Evaluation Department, World Bank.

World Bank. (1980). *Education sector policy paper.* Washington, DC: World Bank.

World Bank. (1999). *Comprehensive development framework: Questions and answers.* World Bank.

World Bank. (2000a). Can Africa claim the 21st century? Washington, DC: World Bank.

World Bank. (2000b). *India-Second Technician Education Project.* Retrieved September 8, 2015, from The World Bank: Projects and Operations: http://documents.worldbank.org/curated/en/2000/06/728964/india-second-technician-education-project

World Bank. (2002). *Maldives-The first and second education and training projects.* Washington, DC: World Bank.

World Bank. (2003a). *Armenia-Education Financing and Management Reform Project.* Retrieved September 8, 2015, from World Bank Projects and Operations: http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2003/06/23/000090341_20030623135136/Rendered/PDF/25989.pdf

World Bank. (2003b). *Senegal-Higher education project.* Washington, DC: World Bank.

World Bank. (2007). *A guide to the World Bank.* World Bank.

World Bank. (2015a). *Beneficiary assessment*. Retrieved August 20, 2015, from World Bank: http://siteresources.worldbank.org/INTPSIA/Resources/490023-1121114603600/beneficiary_assessment.pdf

World Bank. (2015b). *World Development Indicators*. Retrieved 2015, from The World Bank: http://data.worldbank.org/data-catalog/world-development-indicators

World Bank. (2015c). *Worldwide Governance Indicators*. Retrieved 2015, from The World Bank: http://info.worldbank.org/governance/wgi/index.aspx#doc

World Bank (2016a). International Bank for Reconstruction and Development, website. Accessed March 3, 2016 from [http://www.worldbank.org/en/about/what-we-do/brief/ibrd].

World Bank (2016b). International Development Association, website. Accessed March 3, 2016 from [http://www.worldbank.org/ida/what-is-ida.html].

World Bank Independent Evaluation Group. (2009). *2009 Annual review of development effectiveness: Achieving sustainable development.* World Bank Publications.

World Bank Independent Evaluation Group. (2011). *A portfolio note: World Bank support to education since 2001.* Washington DC: The World Bank Group.

World Bank Independent Evaluation Group. (2012). World Bank project performance ratings-codebook.

World Bank Independent Evaluation Group. (n.d.). *Harmonized evaluation criteria for ICR and OED.* Retrieved November 5, 2013, from World Bank Independent Evaluation Group: http://ieg.worldbank.org/Data/HarmonizeEvalCriteria.pdf

World Bank Operations Evaluation Department. (2005). *Improving the World Bank's development effectiveness: What does evaluation show?* Retrieved September 8,

2015, from Open Knowledge Repository: https://openknowledge.worldbank.org/handle/10986/7225

World Bank Operations Policy and Country Services (OPCS). (2006). *Implementation Completion and Results Report: Guidelines.*

Xu, Y. C., & Weller, P. (2009). *Inside the World Bank: Exploding the myth of monolithic Bank.* Palgrave Macmillan.

Zoido, P. (2008). *Public spending on education in Latin America: Does it pay?* Paris: OECD Development Centre.