Evaluation of Developmental Projects of the World Bank in the Western Balkans

By

Albana Rexha

Submitted to

Central European University

Department of Public Policy

In partial fulfillment for the degree of Master of Arts in Public Policy

Supervisor: **Dr. Achim Kemmerling**

CEU eTD Collection

Budapest, Hungary

2013

ABSTRACT

The topic of evaluation in developmental community has gained importance and its usage is increasing. The World Bank is one of the largest multilateral developmental agencies and is a leading organization in evaluation. Therefore, this paper assesses project evaluations of World Bank by analyzing projects implemented in Western Balkan for a period of ten years. Western Balkan represents all types of projects that World Bank implements worldwide by making it an interesting subset to analyze. The purpose of this paper is to find out how informative are these evolution reports in the sense of the output that can be interpreted from them. Thus, having or not having informative reports show also how accountable or not accountable, respectively, is the bank to other stakeholders. The results show that the bank has a high degree of accountability toward other stakeholders like donors and intended beneficiaries. These findings apply particularly to Western Balkan region, but taking into consideration that this region presents a comprehensive range of projects that the bank implements, there is a high likelihood that the same results would be found in other regions.

CEU eTD Collection

This paper is dedicated to *my parents*.

CEU eTD Collection

Table of Contents

CEU eTD Collection

List of Acronyms

Asian Development Bank (ADB)

Center for Global Development (CGD)

Development Assistance Committee (DAC)

Evaluation Cooperation Group (ECG)

European Union (EU)

International Monetary Fund (IMF)

Independent Evaluation Group (IEG)

Monitoring and Evaluation (M&E)

Organization for Economic Co-operation and Development (OECD)

Operations Evaluation Department (OED)

UK Department for International Development (DFID)

Project Implementation Status and Results (ISR)

Project Implementation Status and Completion Results (ISCR)

World Bank (WB, the bank)

CEU eTD Collection

List of Tables

List of Figures

CEU eTD Collection

Introduction

"In my eyes, Americans as well as other tax payers are quite ready to show more generosity. But one must convince them that their generosity will bear fruit, that there will be results."

Paul Wolfowitz,
President, World Bank

The effectiveness of development assistance[1] has been widely discussed and debated among academics and policymakers. Polarized views are held with regard to its impact on society. Scholars such as Easterly (2006) argue that development assistance does not make any difference to the lives of poor people. On the other hand, scholars such as Sachs (2005) try to prove that foreign aid has had an immense impact on reducing global poverty. At the same time, the pressure on aid budgets from tax-payers side has increased over time, they demand accountability (OECD 2010). Uncertainties about the impact of development assistance presented above, and the demand for accountability from tax-payers side, increase the need among the development community to evaluate development assistance.

Assessing whether the objectives of a development activity were met, and how efficiently and effectively they were met entails formal evaluation. Evaluation also serves to find out which development activity was a failure or a success, thus learning the lesson with the purpose of enhancing development assistance. World Bank (WB 2011) defines evaluation as "the process of determining the worth or significance of a development activity, policy or program ….. to determine the relevance of objectives, the efficacy of design and implementation, the efficiency

[1] Development assistance , development aid, foreign aid, refers to the same concept in the whole paper, unless stated otherwise.

CEU eTD Collection

or resource use, and the sustainability of results. An evaluation should (enable) the incorporation of lessons learned into the decision-making process of both partner and donor". WB definition relies on the objectives of evaluation, therefore, this type of evaluation definition is a common definition found in other development agencies, as well. For instance, the Development Assistance Committee (DAC), defines evaluation as "the systematic and objective assessment of an on-going or completed development intervention, its design, implementation and results" (OECD, DAC 2010), which is similar to that of WB. Another key term used throughout this paper is project. A project as opposed to other developmental activities is clearly defined in terms of having specific outcomes as end targets, it covers a smaller geographical area, it always has a start and end date, and is also funded by a specific agency (Bamberger 2009).

The WB[2] is one of the worldwide leading organizations in evaluation (Cracknell 1989). It is also one of the largest multilateral development agencies in the world. The Independent Evaluation Group (IEG) is the internal evaluator body within the structures of the bank that carries out evaluations of all developmental activities including projects. Therefore, I have chosen to assess the WB,IEG report evaluations of projects implemented in Western Balkan region. This region represents a comprehensive range of projects that the bank implements making it an interesting subset that lends itself for analysis. Western Balkan refers to the six South East European Countries including Albania, Bosnia and Herzegovina, Kosovo, Macedonia, Montenegro, and Serbia that still are not part of the European Union (EU). The aim of this paper is to assess how informative are the WB,IEG evaluation project reports to other stakeholder like donors and intended beneficiaries, by looking at implemented projects within a ten year period in Western Balkan region. Informative refers to how much output can actually be interpreted from the data

[2] The bank refers as well to the World Bank (WB) in the whole paper

CEU eTD Collection

available in project evaluation reports. Thus, having or not having informative reports show also how accountable or not accountable, respectively, the bank is to other stakeholders. An important indicator that reveals the degree of accountability is the independency of the IEG. This paper measures output through the number of evaluation reports found, investments per sector, sector performance, the effect of intermediate results indicator in overall project outcome, bank performance and borrower performance for specific projects, and country performance which includes all projects implemented in a country for ten years in comparison to other countries. This paper will present predicted results for each of the outputs based on the literature, and then compare it with the actual results. To sum up, depending on how many of the above mentioned outputs can be derived from reports, shows also how informative are the WB,IEG project evaluation reports.

This paper has three chapters. The first discusses the theoretical framework of evaluating development assistance. It reviews well-known articles and books that discuss the purposes, and rationalities of conducting evaluations. It also discusses the challenges of evaluating development aid. The second provides a brief history and recent development of evaluations within the bank. The core part of this chapter includes the information about the work of the IEG, the responsible body for evaluations within the bank and outlines its procedures for evaluating projects which will also be used in data analysis. The third presents the data gathering, analysis, and results from project evaluation reports. Based on these analysis this paper will provide answers to the question of how informative are the WB,IEG project evaluation reports to other stakeholders.

CEU eTD Collection

Methodology

To answer the question how informative the WB project evaluation reports are, this paper uses secondary data from the online database named "Projects and Operations" of the WB group (2013). Based on the data gathered from the bank database, I created my own dataset with all variables necessary and available to answer this paper's research question. The dataset includes specific information about each project including the project name, country, sectors of investment, commitment amount, evaluation report results, approval date, closing date, bank performance, borrower performance, overall project performance, intermediate result indicators.

Furthermore, the dataset created covers 109 closed projects implemented in the Western Balkan region for a period of ten years, 1 January 2003 until 31 December 2012. A closed project refers to any project that was finished within the ten year period. I have chosen to work with a ten year period for the analysis because I think it is more representative (rarely things happen by chance for a period of ten years). The analysis will also provide a more insightful picture of how informative are project evaluation reports because a longer period and a larger number of reports are assessed. This paper uses two programs, EXCEL and Microsoft SQL Server 2008 express edition to conduct data-mining with the purpose of extracting information in a more structural and understandable way. Examples include charts, tables, and graphs. Through these two programs the secondary data serves as the basis of quantitative analysis through which this paper checks if the predicted results based on literature match the actual results, thus providing an answer to the question of how informative the bank evaluation report projects are to other stakeholders. The following chapter provides the theoretical framework.

CEU eTD Collection

CHAPTER I. Theoretical Framework

This chapter will discuss different purposes of conducting evaluation in the development field. It will also discuss methods of evaluation and the most prominent criteria of evaluating development assistance. The last part will review challenges of evaluating developmental activities.

1.1 The Rationale, and the Purposes of Evaluation in the Development Field

The beginning of development assistance can be traced back to the 60s (Pacquement 2010) and is mainly discussed in terms of allocation and effectiveness. For decades, an intense debate has been going on among academics and practitioners about the effectiveness of aid in creating economic development and alleviating poverty. A common question found in the literature is - does aid work (Riddell 2007). Polarized views are held with regard to the answer of this question. Scholars such as William Easterly (2006) argue that aid has failed to show any positive impact on development in third world countries. A similar but more heretical view holds the economist, Dambisa Moyo, who in her book "Dead Aid" (2009) calls for termination of aid toward African countries. However, on the other hand Jeffrey Sachs (2005) invites rich governments of the world to give 0.7% of their GNP in the form of aid to poor countries in order to end poverty. An in between view is held by scholars like Collier & Dollar (1998), Alesina & Dollar (2000) who argue that aid should be given to countries that have established sound institutions, and that this is the only way that aid could have positive impacts in development. In view of these polarized opinions about aid effectiveness, pressure on aid agencies to be more accountable to tax payers has increased over time (OECD 2010). Donors (tax-payers) want to see

CEU eTD Collection

'measurable results'; they want to see if their money made a difference in society (Hailey et al. 2005). Thus, donor community is compelled to show the impacts of their activities on societies through the process called *evaluation*. Evaluation serves as a means to assess donors impact in society and tries to make development agencies more accountable to its donors-citizens (Cracknell 2000).

Evaluation is defined by academics and donor organizations in similar ways. For example, the well-known scholar in the field of evaluations, Michael Scriven, defines evaluation as "a process of determining merit, worth, or significance" (2007). CARE a leading worldwide organization in fighting poverty defines evaluation as "the periodic assessment, analysis and use of data about a project" (Barton 1997). The Center for Global Development (CGD) (2006) views evaluation as a means of addressing a 'gap' in the development field. CGD refers to this 'gap' as the data that could improve the effectiveness of aid. From definitions framed within donor organizations and by academics, it can be implied that the purpose of evaluation is about improving developmental outcomes while gathering data periodically about developmental activities like projects and programs. However, before providing conclusions for the rational of conducting evaluation from the definitions, this paper first addresses the literature on different purposes of evaluating development aid.

Scholarly literature provides different purposes of evaluating development assistance. For example, Cracknell in his comprehensive study "Evaluating Development Aid: Issues, Problems, and Solutions" (2000), recognized evaluation as one of the most growing fields in the world, covering all activities that a donor agency performs. He summarizes the rationale of evaluating

CEU eTD Collection

development aid in two main purposes 'accountability' and 'lessons learned'. In the context of aid community 'accountability' is a means of reporting your activities - the way you spend the money, to a higher authority (Crawford et al. 2003). 'Lessons learned' has to do with identifying why certain activities failed and some others succeeded in reaching their objectives, thereby ensuring that aid community learns the lessons that could be useful for future activities. Overall the purposes of evaluation have to do with *proving* that changes are taking place, and *improving* future aid interventions.

Palumbo and Nachmias (1983) also emphasize the role of evaluations in decision making. For instance, if the development agency is building schools, through evaluation they can know whether it is worthwhile to continue building schools or is it better to switch over to roads or hospitals. They group the purposes of developmental evaluation into four categories - 'terminate-continue', 'program improvements', 'informational component in decision-making', 'inform and educate society'. The first two categories have to do with improvements from failures or successes of activities, which is also similar to the 'lessons learned' as described by Cracknell. Evaluation is also conducted for purposes of ensuring efficiency, efficacy, and effectiveness on developmental activities (Checkland 2001, Crawford et al. 2004). Efficiency refers to achieving higher ratio of outputs with a certain ratio of inputs. Effectiveness refers to the achievement of the goals set at the beginning of the activity. Efficacy is similar to effectiveness but under ideal/controlled conditions. To sum up, evaluation serves as a tool of gathering information about an on-going or finished project, in order to show its impact on society while portraying measurable results to a higher authority. Thus, evaluation is conducted for a variety of purposes including accountability, lessons learned, and decision- making. Following the discussions about

CEU eTD Collection

the role of evaluation in developmental field, the next section will address methods and types of conducting evaluation.

1.2 Types and Methods of Evaluation

Each project is different in terms of time-frame, data available for the project, impact on society, costs, and the like. Therefore, the ways one could evaluate a project also varies a lot. A broader division of evaluations is made between formative and summative types of evaluations. The evaluation expert, Scriven (1967) was one of the first authors to distinguish between formative and summative evaluations in education programs. According to Scriven, formative evaluations are used on the early stages of project development, while summative evaluations are conducted at the final phase of project life, in order to see the "final product". Formative evaluation also referred to as monitoring involves analysis of inputs, activities, and outputs. Summative evaluations consists of outcomes and impacts in society (Asian Development Bank 2006). Thus, monitoring is considered part of evaluation, and one spectrum interprets them as processes that go hand in hand with each other. Particularly, in the article of Crawford and Bryce (2003), monitoring and evaluation (M&E) are referred to as "intimately linked".

UNDP defines M&E as two processes that "differ but are closely related" (1997). According to these perceptions monitoring differs from evaluations, still it is a part of it. Monitoring differs since it mainly discusses the outputs, whereas evaluation is about outcomes. Multilateral organizations such as Asian Development Bank refers to monitoring as a "traditional" part of evaluation which looks only at outputs, whereas outcomes and impact evaluation refers to a more modern type of evaluation which looks at development results (2006). This closeness and

CEU eTD Collection

division of M&E is shown on fig.1. Closeness in terms of the main or final purpose-accountability to the donor, and difference in terms of output and outcome.

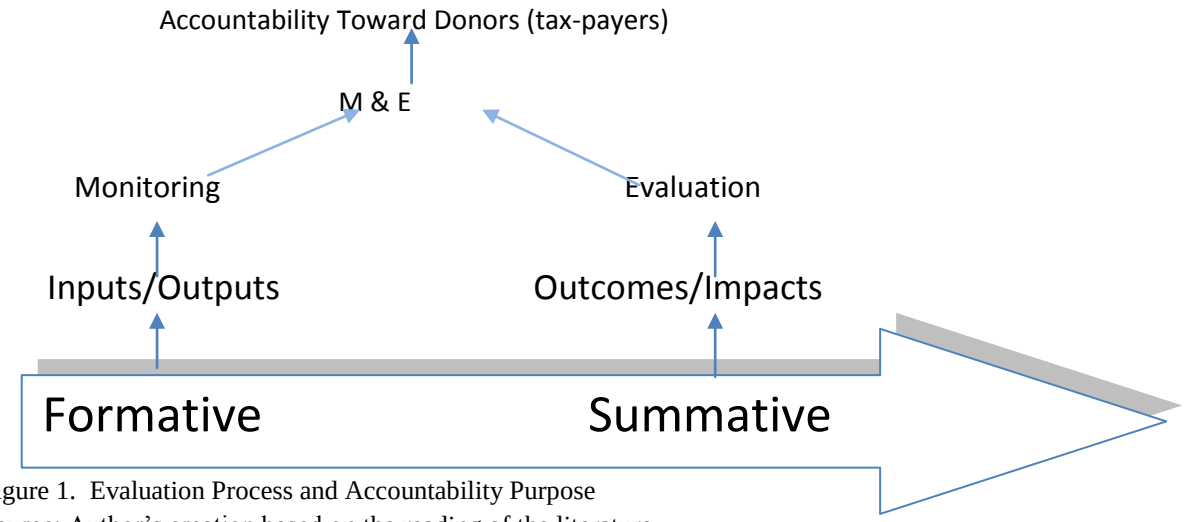


Figure 1. Evaluation Process and Accountability Purpose
Source: Author's creation based on the reading of the literature
Note: The arrow only shows beginning point of the evaluation

Inputs and outputs as seen in fig.1 are known as parts of the monitoring process which is an on-going process during the life of the project. Inputs are resources that are employed at the beginning of the project including staff, money, time, and equipment. Output are the services or the products that are delivered to the recipient at the end of the project. Examples of output could be the number of trained people, the number of grants given and the like. This is known as the formative part of the evaluation process and all the information gathered during this process is important for the summative evaluation which includes outcomes and impacts on society. Outcomes are measured in "non-monetary quantitative terms" (World Bank 2004). An example of an outcome could be an improvement in student reading scores. Through outcomes we can better understand the impact on final results, because impact shows the extent to which the well-being of people has been improved because of a development activity intervention. Both terms

CEU eTD Collection

'outcome' and 'impact' are very difficult to distinguish, because they both tend to show the "degrees of influence" of an aid intervention in society (Naonobu et al. 2009). The more effort one puts to the formative evaluation, the less resources are needed in the summative evaluation .

Impact evaluations can be of quantitative and qualitative nature (Asian Development Bank 2006, Baker 2000). Quantitative approach can be experimental or randomized and non-experimental or quasi-experimental methods. For the purposes of conducting quantitative impact evaluation in either way, experimental or non-experimental, there is a need for two groups, control and intervention group. The intervention group refers to the group where aid intervention took place, whereas the control group refers to the group where no intervention takes place (World Bank 2006). Then, the intervention group results are compared with the control group to check for the possible outcomes. Baker (2002) discusses both approaches. First, the experimental or randomized design argues that control groups are randomly selected. This means that each participant has an equal chance of being selected and expectations for results are the same in both groups, the control and the intervention group. Second, the non-experimental or quasi experimental design portrays the complete opposite of the randomization. Through this approach the development agency tries to find the control groups that are almost identical but different in one variable that is aimed to be measured, within the project group. On the other hand, the qualitative approach designs tend to show the opinions of intended beneficiaries through questionnaires. This kind of evaluations show how the development activity has had an impact on recipients according to themselves (Mohr 1995).

Furthermore, evaluation could be conducted by an internal or external body. According to Weiss (1998), both types has their own advantages and disadvantages with regard to "administrative

CEU eTD Collection

confidence", "objectivity", "understanding of the program", and "autonomy". Weiss elaborated these points by concluding that objectivity and autonomy are advantageous for external bodies or auditors. On the other hand, understanding of the program is advantageous for internal bodies who are familiar with the programs and projects. The last point which does not clearly point up any of the two kind of bodies is administrative confidence. Weiss, claims that this point has its own advantages for both bodies by claiming that internal evaluators can be more practical as opposed to external, but internal evaluators are viewed as less credential when compared to external auditors who have more reputation. Following the discussion about the process and ways of conducting evaluation, the next section will address the differences between two main types of developmental activities, projects and programs because they are often confused by readers in general.

1.3 Developmental Activities: Projects and Programs

The word development activity can refer to any social or economic project or program financed by bilateral, and multilateral development agencies. Multilateral aid agencies are international organizations such as the International Monetary Fund (IMF), the World Bank (WB), etc (Labs 1997). Countries themselves providing aid are known as bilateral or national donors (Easterly 2007) such as the UK Department for International Development (DFID). It is important to note the difference between a project and a program, albeit both are activities conducted by the development community. "Though some overlap exists, projects and programs are distinctly different, requiring practitioners who audit them to understand what separates the two" (Marinaccio et al. 2012). Bamberger (2009), defines programs as broader interventions, consisting of several projects. Broader intervention means that programs cover larger

CEU eTD Collection

geographical areas as opposed to projects. A project also refers to a more specific intervention which is concerned about reaching certain outputs (Marinaccio et al. 2012). To sum up, a project is different in that is usually shorter in time, is focused in specific outputs, is more clearly defined, has a start and end date. Following the discussions about the difference between projects and programs, the next session will discuss few challenges faced by all kinds of developmental activities like projects and programs.

1.4 Challenges of Evaluating Developmental Assistance

Scholarly literature discusses three main challenges faced by development community when conducting evaluation for all kinds of developmental activities. These challenges include the involvement of politics, the high costs, and the difficulty to find the right indicators that measure the changes in certain projects such as capacity building ones (Chelimsky 1987, Weiss 1973, Baker 2000, Cracknell 2000, Crawford et al. 2003). First, the history of politics and evaluation is almost as old as evaluation itself, and it boomed during 1975-1980 when a set of articles and studies were published (Chelimsky 1987). “A theory of evaluation must be as much a theory of political interaction as it is a theory of how to determine facts” states Lee Cronbach (1980) (quoted on Chelimsky 1987). According to this quotation evaluation shows a better way of convincing others through results which are measured through certain indicators set in the influence of politics. Another way that politics gets involved in evaluations is through aid interventions providing social and economic changes which “attract” a range of stakeholders who want to see and show results to tax payers (Crawford et al. 2003). All this process is known as public policy, and as noted in the Chelimsky (1987) article, public policy is a ”product of politics”. For instance, randomization as a method of quantitative approach type of evaluation, is

CEU eTD Collection

often considered "politically unfeasible" (Grun 2006). "Political unfeasibility" refers to the high risk of choosing randomly participants to be part of aid interventions without being aware of any possible predicted results, because by randomly choosing one can include a participant that might have low outcomes. None of the developmental agencies would want to have low outcomes, because this might negatively affect their reputation. This also presents one of the rationales of politics being a part of evaluation for such a long time. This is also proven by the existence of a much smaller number of randomized evaluations as opposed to the non-randomized number of evaluations (Pitman et al. 2005). Among the few randomized evaluations it is important to note one of the most well-known evaluations - the Progresa - a designed program to help improve school participation in Mexico (Skoufias et al. 2001). Taking into consideration the rationale for the involvement of politics in evaluation, it seems that politics will continue to remain a challenge of evaluating development aid.

Secondly, conducting evaluation requires the use of extra financial resources (ADB 2006), thereby making it expensive for small –scale development agencies, or sometimes even for the bigger ones. This becomes a greater challenge considering the fact that donor agencies are mainly the ones who pay for the evaluation, thus influencing considerably the indicators chosen to evaluate the developmental activity (Bamberger 2000). Baker (2000), argues that for a well-designed evaluation there is a need for expert evaluators and data collection which raises the cost of evaluation. As discussed above one of the ways of doing impact evaluations is also through qualitative approaches that consist of questionnaires. These questionnaires involve high costs of preparation, and implementation (Crawford et al.2004).

CEU eTD Collection

The third challenge concerns a group of developmental activities that are more complicated to measure due to their less tangible processes. Projects such as organizational capacity building tend to be described as "ill-defined" processes which present another evaluation hurdle for development community (Hailey et al. 2005). It is also important to note that if one wants to measure all changes in these kinds of projects while putting everything in measurable indicators, the risk of not capturing all changes increases (Taylor 2003). This way the tendency of false measurement increases, since it is not easy to translate into quantitative data the capacity building projects and to capture all the social and economic changes. The USAID, has also claimed that capacity building projects can be of "subjective nature", hence provide "false" measurements of developmental results (USAID 2000). To sum up, all these hurdles present major challenges to the evaluating developmental community while also making evaluation more costly in terms of time and finance. Until now this paper has addressed the objectives, the rationales, the types, and the challenges of evaluating developmental activities but now it is important to review the criteria for conducting evaluations (what criteria does the development community use in general) and I decided to discuss DAC criteria, because they present the most prominent evaluation criteria, thereby the next section addresses these criteria.

1.5 The Most Prominent Criteria for Evaluation: OECD, DAC

Evaluation is becoming a crucial and a widely used tool in the development community but it is important to note that the developmental evaluation is not a 'fashion' rather it is a demand of other stakeholders like tax-payers (Naonobu et al. 2009). These tax-payers want to see if their taxes are being spent efficiently, and if that is making any difference in the life of others, because that portrays the basic purpose behind all the money spend by development agencies. Therefore,

CEU eTD Collection

the Development Assistance Committee (DAC), part of the OECD, has been committed to strengthening the developmental effectiveness aid since 1991 when it first came up with five main evaluation criteria (OECD 1991). The criteria of 1991 report included effectiveness, efficiency, impact, relevance, and sustainability. Part of the report discussed the involvement of other stakeholders which represents the intended beneficiaries and also those indirectly affected by the aid intervention, in the evaluation process. The involvement of new stakeholders in evaluation is considered a new trend, because before evaluations were conducted only by the donor community. The benefit of engaging other stakeholders in the evaluation process has to do with being aware of new perspectives about aid impacts and outcomes about developmental activities be it projects or programs. The DAC criteria are the most prominent and since 1991 have been permeating the development community (Chianca 2008). Most bilateral, and multilateral development agencies have adopted their evaluation framework based on the DAC criteria. The prominent use of these criteria is explained through the composition of DAC members representing the thirty strongest democracies worldwide. Therefore the importance of these criteria toward enhancing aid evaluation has been crucial.

This chapter provided an insightful picture of evaluating development aid by addressing the purposes, the types and methods, the challenges, and the most prominent DAC evaluation criteria. Therefore, the next chapter will discuss in particular the World Bank procedures for evaluating projects.

CEU eTD Collection

CHAPTER II. World Bank Evaluation Procedures

This chapter will provide a brief background about aid evaluation in the WB, because the long history starting in the 70s of evaluation function in WB also shows the relevancy of assessing WB evaluation projects in this paper. The second part will discuss the procedures followed by the Independent Evaluation Group (IEG), the independent internal body responsible for aid evaluation within WB. Particularly, addressing the methods and ways of project report evaluations, because the analysis in the third chapter will assess project report evaluations while mainly taking into consideration procedures discussed in this chapter.

2.1 Background of the Evaluation Function in the World Bank

WB is known as a leader in the field of aid evaluation because it has been spending a lot of resources toward development assistance evaluation, and has tried to encourage the improvement of evaluation in developing countries (Cracknell 1989). During the 70s, the former president of WB, McNamara established the evaluation function in order to measure developments and the contribution of WB toward enhancing people's lives (Grasso et al. 2003). McNamara initiative is considered the first step in the history of WB toward the establishment of evaluation and later on the basis for the creation of Operations Evaluation Department (OED). Thus, it started as a division, then it became a unit, and after another three years, in 1973, the unit become a department in order to avoid conflicts of interest (WB archives). OED as a bank department within years gained even more independence by choosing its own director, and by reporting directly to the board of directors (Grasso et al. 2003). Reporting directly to the board leaves no room or at least much less room for influence by directors of developmental activities in

CEU eTD Collection

evaluations. McNamara, also noted lessons learned as another important purpose of establishing evaluation function within WB. He noted that it is important to distinguish between failure and successful developmental activities, in order to learn from them. In 2005 the official name of OED was changed to the Independent Evaluation Group (IEG) which stands today, as well. IEG is responsible for also evaluating the International Bank for Reconstruction and Development (IBRD), the International Development Association (IDA) and the work of the International Finance Corporation (IFC), the Multilateral Investment Guarantee Agency (MIGA), (IEG, About us). Following the brief history of the establishment of evaluation function and the IEG, the next section will address the work of IEG because in the third chapter this paper will analyze evaluations prepared by this body. Thus, it is important to know how IEG works.

2.2 The Independent Evaluation Group (IEG)

IEG is the only independent body in WB structures in charge of evaluating developmental activities including programs, projects, and policies. It has its own general director and reports directly to the WB Group's Board of Directors. The purpose of lessons learned while finding out which project worked and which did not work, has remained one of the crucial purposes since IEG's establishment. It assesses the "relevance", "efficiency", and "efficacy" of WB developmental activities in order to measure the contribution of WB in achieving greater developmental effectiveness (IEG 2011). In 2011, IEG conducted a self-evaluation while assessing its own performance as opposed to other best practice evaluation of developmental agencies. Evaluating yourself, just as IEG did by evaluating themselves or otherwise referred to as *evaluating the evaluator*, is an important sign of organizational maturity, and of accepting that not all developmental activities conducted by WB are highly successful. This also shows

CEU eTD Collection

willingness to learn from mistakes, and to improve in future aid interventions. The self-evaluation was conducted by a group of external auditors and staff members. They assessed the work quality of IEG based on Evaluation Cooperation Group (ECG) good practice standards and conducted surveys with staff members and other stakeholders (IEG 2011). The self-evaluation did not evaluate any specific project that IEG evaluated during years. But, it was a more broader evaluation which ended up with recommendations on how to keep the independency of IEG (IEG 2011), because being an internal evaluator based on literature (Weiss 1998) has the disadvantage of easily being biased toward the organization, in this case toward the WB. But, according to the self-evaluation, they are a complete independent body reporting directly to the WB Board of Director.

However, it is not possible to assess the methodologies that IEG uses for evaluation purposes because it does not share any information on its web-page. Although the disclosure of information would show a high level of transparency to other stakeholders including tax-payers and intended beneficiaries. It would also be helpful for scholars who want to analyze the methodologies that IEG uses. But there are two ways one could get to learn more about the work of IEG in evaluating developmental activities. First, IEG held the six in row "Evaluation Week" with the goal of sharing lessons learned and best practice with other developmental agencies (IEG 2013). Other scholars like Tuan (2012), also claim that WB makes it possible to spread lessons learned from evaluations through these seminars. Sharing best practices is considered a 'public good' that benefits everyone beyond national borders (Pitman et al.2005). Second, IEG shares online its evaluation reports like project implementation and completion results report reviews, project performance assessment reports, working papers and the like (IEG, access to

CEU eTD Collection

information). Uploading these reports online is important, and to my knowledge this is the only possible way one could get more information about the evaluation procedures since the section on methodologies does not have any document written. Therefore, the following section will derive procedural evaluation information from reports with the purpose of using them in the analysis part in the third chapter.

2.3 Procedural Information about Project Evaluations in the World Bank

As discussed above the IEG conducts evaluations to a variety of aid interventions. Project is one type of development activity that IEG evaluates and which this paper assess in more detail by looking at projects implemented in Western Balkan within a ten year period. A project refers to "clearly defined" , "time bound interventions", with a "defined funding source " (Bamberger 2009). According to IEG, the WB Project Performance Rating database is the oldest database of its kind that includes about 8000 project evaluations since its establishment (WB Rating). IEG provides two types of reports for project evaluations including Project Implementation Status and Results (ISR), and Project Implementation Status and Completion Results (ISCR). First, the ISR is a standardized, four page document that includes basic information about the project such as title, country, status, sector, approval date, closing date, and commitment amount. The ISR also summarizes main objectives of the project, and offers a yes or no answer on the achievement of objectives (PISR, 2012). Another important part of the ISR are the "results indicators". Each result indicator in the ISR has the name, the "unit of measure", the "baseline", the "current", and the "end target". "Unit of measure" could be different depending on the indicator itself. For example, the indicator could be measured in days, or it could be a simple yes or no answer. The "baseline" measures the situation before the aid intervention take places and is used to compare

CEU eTD Collection

with the end target in order to see the outcomes. The "current" data is measured during the project life and shows the situation by the time project is measured. The "end target" is measured after the project is finished and in a way shows if the objectives are achieved. The ISR gives an insightful but brief evaluation picture of a project.

Second, the ISCR is a longer report as opposed to ISR and provides more information about the project life and end results. It is not a standardized form in terms of page number or the amount of information per section, albeit it has similar sections in each evaluation report including the basic info about the project, a summary, the relevance and outcomes. In the beginning it was only the total project outcomes that IEG evaluated but within years, it enhanced performance ratings and included bank and borrower performance, as well (WB Ratings). In most of the cases, an ISCR table of content looks like the one in fig.2.

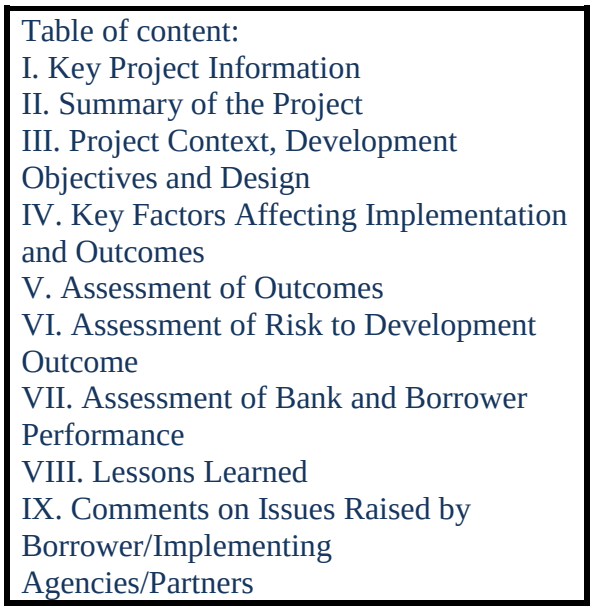


Figure 2. Author's creation based on other projects table of content

It is important to review what project outcome, bank and borrower performances mean, and how are they measured since these terms and measurements will be used in the analysis in chapter three.

CEU eTD Collection

- *Project outcome performance* is a rating that "captures the extent to which a project's major relevant objectives were achieved or are expected to be achieved, efficiently" (IEG 2011). The word efficiently, and relevancy shows that this rating captures important criteria for evaluation.
- *Bank performance* evaluates the quality of services the WB provides to its beneficiaries, at the beginning and throughout the project life cycle (OED 2005).
- *Borrower performance* is the third rating which evaluates the performance of the beneficiary. OED (2005) defines this rating as the degree to which the borrower shows willingness and ability to guarantee and comply with the criteria and requirements agreed with the bank.

The three performances are rated using the same six scales including highly satisfactory, satisfactory, moderately satisfactory, moderately unsatisfactory, unsatisfactory, and highly satisfactory, as shown below in tab.1 .

Table .1 IEG Performance Rating Scales

	Scales	Explanation
1	Highly Satisfactory	***No*** **shortcomings in identification, preparation, or appraisal**
2	Satisfactory	***Minor*** **shortcomings in identification, preparation, or appraisal**
3	Moderately Satisfactory	***Moderate*** **shortcomings in identification, preparation, or appraisal**
4	Moderately Unsatisfactory	***Significant*** **shortcomings in identification, preparation, or appraisal**
5	Unsatisfactory	***Major*** **shortcomings in identification, preparation, or appraisal**
6	Highly Unsatisfactory	***Severe*** **shortcomings in identification, preparation, or appraisal**

Source: Author's table based on OED 2005 report

CEU eTD Collection

The difference is clear among some of the scales, for instance the first one with no shortcomings and the last one with severe shortcoming. However, the difference between moderate and significant is not quite clear. There is no definition in the report that specifies what exactly moderate means and what is the difference from significant that made them specific categories. With this limitation in mind, these scales according to the explanatory statements in tab.1, will be used in chapter three, since all projects are evaluated according to this scale.

To sum up, IEG provides two types of project evaluation reports ISR and ISCR. The ISR one is shorter and has a standardized form, whereas the ISCR is longer and a more detailed one. The analysis in chapter three has mainly looked at the ISCR because more information such as bank, and borrower performance could be seen. Following the discussions about IEG as the internal independent body responsible for WB developmental activity evaluations, and some procedures particularly for project evaluations that could be derived from report analysis, the next chapter will address the results based on quantitative approach data analysis.

CEU eTD Collection

CHAPTER III. Data Analysis and Results

This chapter presents the secondary data gathered and the quantitative analysis conducted on the basis of these data. The data analysis and results are categorized in two groups. Data gathering in the first section which is shorter shows that not all projects are evaluated, thereby it tends to identify whether there are visible and identifiable criteria that makes a project more likely to be evaluated as opposed to another. The second group gathers the same data for each project including commitment amount, sectors of investments, number of projects and the like for the purpose of analysis and comparison. The data are analyzed in more detail to find out how informative evaluation reports are in the sense of how much output can actually be interpreted. A brief summary of findings will follow.

3.1 Description of the Data

The secondary data derived from the WB dataset on "Projects and Operations" includes 109 implemented projects in the Western Balkan for a period of ten years. The dataset of the bank contained a variety of information for projects, however, I have gathered only the data needed to answer the research question. Thus, in the data-base I created, each project contains the same information for the purposes of analyzing and comparing the data. The information includes the project name, country, approval date, closing date, commitment amount, % spend in each sector for each project (usually a project is spread into several sectors), intermediate results indicator which are conducted while the project is running, implementation completion results report which is conducted at the end phase of the project cycle, project outcome performance, bank and borrower performance. The data gathered are approached in a quantitative manner using EXCEL

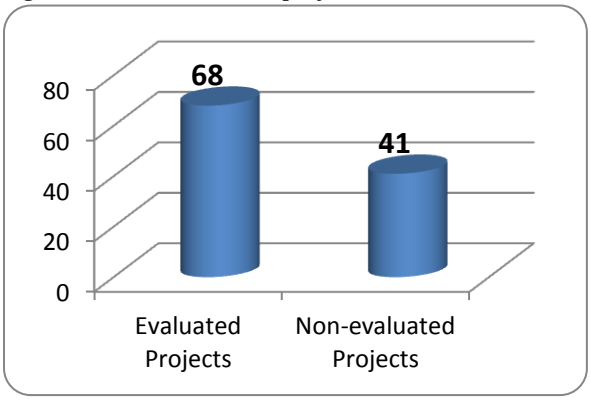
CEU eTD Collection

and Microsoft SQL Server 2008 express edition, and are provided in a more structured way through graphs, and tables. I run data-mining to analyze the data using ex-ante rationality while comparing my predicted results based on literature with the actual results.

3.2 Analysis and Results of all Projects

Data gathering and calculation shows that the bank implemented 109 projects in the region of Western Balkans for a period of ten years. Projects were implemented in different sectors including economy, law, security, public administration, trade, IT, telecommunications, agriculture, banking, finance, health, education, mining, energy and the like. The amount spend on each project differs ranging from the lowest 0.1 to the highest 111 million. The data also show that not all implemented projects implemented within ten years were evaluated. As shown in fig.3, 68 out of 109 projects have an evaluation report.

Figure 3. Total number of projects evaluated and non-evaluated



Source: Author's creation based on the WB data

However, there is no evidence showing why some projects are evaluated and some others are not. Thus, before looking at evaluated projects it is necessary to identify if possible what makes some projects get evaluated and others not. This section tends to identify whether there are visible and identifiable criteria that makes a project more likely to be evaluated as opposed to

CEU eTD Collection

another. For the purpose of finding out this paper checks for any possible relation of the evaluated projects with project commitment amount and intermediate results indicator. The predicted results argue that projects with higher amounts are evaluated compared to those with lower amount of money invested. First, in order to see if there is any relation between the commitment amount and the decision to evaluate or not a project, I have looked at the min and max commitment amounts to a project that has been evaluated and not evaluated. Tab 2. shows that the min value of an evaluated project was $0 while the max was $111 million. The min value of a non-evaluated is $1.5 million while the max is $85 million.

Table 2. Min & Max values of an evaluated and non-evaluated project

	MIN value	MAX value
Evaluation conducted	0	111
Evaluation not conducted	1.5	85

Note: all values are in mil $
Source: Author's calculation in the dataset created based on the WB data-base

The actual results on tab.2 show that there is no relation between the decision to evaluate or not a project and the commitment amount. Second, this paper checks the possible effect of intermediate results indicator in having the final evaluation report. Intermediate indicators are conducted in the mid-phase of the project, and if a project has these indicators it means that this project has been looked upon before. It makes the project look more important than others, because more effort has been spend in that particular project compared to others who do not have these indicators. This increases the possibility for a project that has intermediate indicators to also have an evaluation report. For the purpose of finding out if that is true, this paper checks out the number of times when both intermediate results indicator and evaluation report were present. Then, it checks out the number of times when intermediate indicators where present but

CEU eTD Collection

evaluation report was missing. It also, checks out the number of times when intermediate indicators were missing but the final report was present and the fourth combination checks out the number of times when both intermediate indicators and evaluation report were missing. These four combinations are present in the below tab.3

Table 3. The relation of intermediate results indicator and final evaluation report

Intermediate Results Indicator (IRI)	Implementation Status and Results Report (ISRS)/evaluation report	%
Yes-37	Yes-68	54.44 %
No-31	Yes-68	45.58%
Yes-3	No-41	7.32%
No-38	No-41	98.62%

Source: Author's calculation in the dataset created based on the WB data-base

Results in tab.3 show that chances of having the ISRS (evaluation report) when IRI are present are higher for 8.86% as compared to the opposite case when IRI are missing. So, there is 54.44% chances of having the evaluation report if IRI is present, and there is 45.58% chance of having the evaluation report if IRI is missing. On the other hand, there are low chances of 7.32% of not having evaluation report if IRI is present, and 98.62% chances of not having evaluation report when IRI is also not present. These results show a relation between IRI and evaluation report, however this relation does not assure us that once we have IRI we must have the evolution report. Although, chances of having the evaluation report are higher when IRI is present, and chances of having IRI when the evaluation report is missing are very low. The predicted results in this case are met to a certain extent, but not fully, because as said above it is not 100% reassuring that when we have one we have the other. To sum up, there is no relation between commitment amount and the decision to evaluate or not a project, neither we definitely have an

CEU eTD Collection

evaluation report if we have IRI, albeit the likelihood of having the report is higher when we have IRI. To sum up, the available data analysis show that evaluated projects range within different commitment amounts proving no relation between that and the decision to evaluate or not a project. But, the data shows a higher likelihood for a project to get evaluated if that project has intermediate results indicator. At this point of analysis, there is no possibility (no data available) that would allow us to trace back further the decision for evaluating or not evaluating a project.

Furthermore, going back to the beginning where this section started to assess the projects while calculating the overall number of evaluated and non-evaluated projects, the data gathering and calculation shows also the number of evaluated and non-evaluated projects in each country of Western Balkan, tab.4.

Table 4. Number of evaluated and non-evaluated projects in each Western Balkan country

	Albania	Bosnia and Herzegovina	Macedonia	Montenegro	Kosovo	Serbia
Evaluated Projects	16	9	14	6	10	13
Non-evaluated projects	8	11	2	4	5	11

Source: Author's calculation

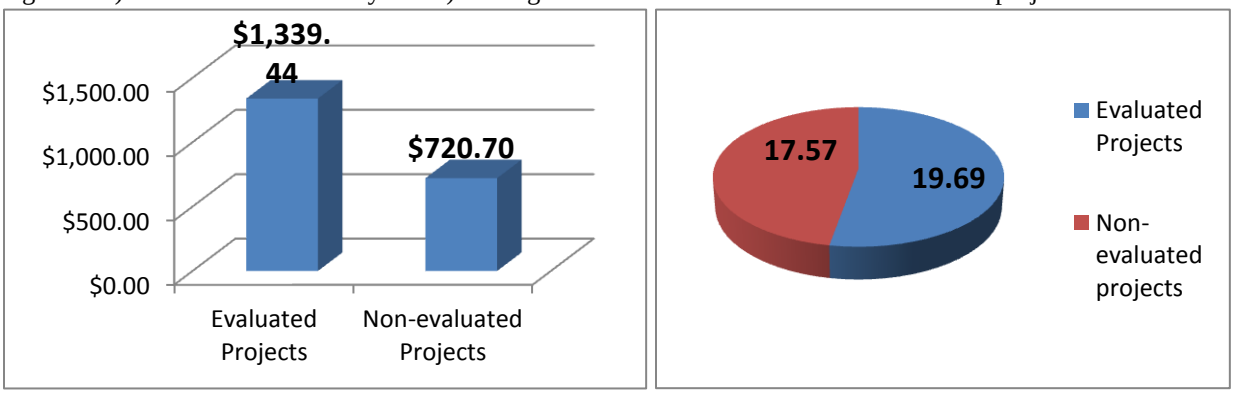
Results show that in each country the number of evaluated projects is higher as opposed to the number of non-evaluated projects. This does not show any bias to any of the countries because all have similar numbers of evaluated and non-evaluated projects, compared to the total number

CEU eTD Collection

of projects implemented in that particular country. This also corresponds with the overall number of evaluated projects which is higher compared to non-evaluated ones.

Another important output that this paper can interpret from the evaluation reports is the total amount of money spent and the average amount of money spent in evaluated and non-evaluated projects. This paper predicts that evaluated projects involve higher amounts of money spend compared to non-evaluated projects. This prediction is based on the number of evaluated projects which is also higher (see fig.1).

Figure 4. a) Total amount of money and b) Average invested in evaluated and non-evaluated projects



Source: Author's creation based on data gathered from WB

This results on fig.4.a show, the total amount of evaluated projects exceeds the non-evaluated total amount. Fig.4.b also shows that the average amount of evaluated projects exceeds that of non-evaluated projects. Looking at the number of the evaluated and non-evaluated projects (fig.1) one might conclude that IEG evaluates more projects, and does not evaluate less projects. However, when looking at the total amount and the average in fig.4 a) and b) there is no such a big difference as in the number of projects. Thus, when looking only at the numbers it might be misleading to conclude that IEG evaluates more or less projects, without looking at the amounts

CEU eTD Collection

spend in each group. This also shows an important output that could be interpreted from evaluation reports after certain calculations and analysis.

3.3 Analysis and Results of Evaluated Projects

This section assess the 68 projects that have evaluation reports with detailed information as opposed to non-evaluated projects which have only specific information available such as commitment amount. Apart from general data including country, commitment amount, the information found in the evaluation reports consists also of data such as outcome performance, bank performance, borrower performance, intermediate results indicator, and sector of investment. This section assess reports by trying to interpret four types of output. First, it compares borrower, bank, and project outcome performance with the purpose of finding out if the bank is rated much higher which according to the literature would make the IEG in the bank not highly independent. Not being independent would also show a low degree of accountability to stakeholders. Second, it assess expenditures per sector measured in $ and as % of the overall amount spend to find out if the three first sectors match the bank objectives. Then, it also looks at sector performance to check if other smaller sectors in terms of expenditures are rated higher as compared to the three highest sectors. Third, it assess the possible positive effect of intermediate results indicator in the project outcome performance. The fourth output looked at to interpret from the data is country performance to find out if best performers based on other indicators such as WB Doing Business are also best rated by WB,IEG.

- Project Outcome, Bank, and Borrower Performance in Comparison

Explanatory statements for each rating, the project outcome, the bank, and the borrower were provided and discussed in more detail in chapter two as part of WB evaluation procedures. Each

CEU eTD Collection

of these ratings was measured on a scale ranging from the highest performance named "highly satisfactory" to the lowest performance named " highly unsatisfactory (see fig.2 for explanatory statements of each scale). The availability of each rating performance gives us the opportunity to find out who performed better according to IEG. This presents important output to interpret from these evaluations, since these results will show a lot about the independency of IEG, as well. As literature addressed in chapter one (Weiss 1998), claims that internal evolution bodies such as IEG, tend to be less objective. Therefore, based on the literature, the predicted results argue that the bank will be the best performer. With the purpose of finding out if that is true, this paper needs to calculate the scales for each rating. But, it is not possible to add up the current scaling and show which performed better. For instance, if we add highly satisfactory and moderately satisfactory no results will be derived. Therefore, I assigned numerical values to each scale. As shown in tab.5. the numerical values range from the lowest-1 to the highest-5.

Table 5. Values assigned to each scale

Scales	Numbers assigned
Highly Satisfactory	5
Satisfactory	4
Moderately Satisfactory	3
Moderately Unsatisfactory	2
Unsatisfactory	1

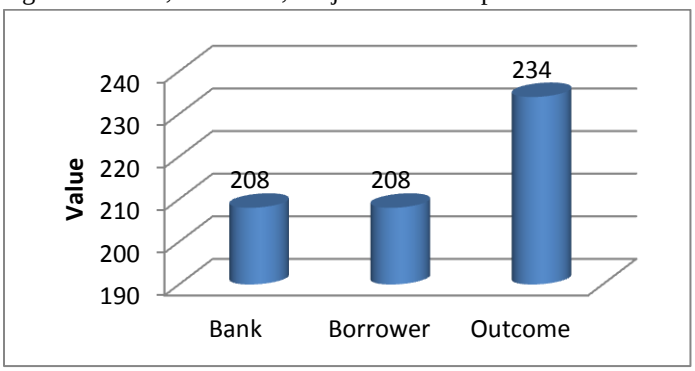
Source: Author's creation with the purposes of analyzing and comparing three performance indicators

I excluded the 'highly unsatisfactory' scale because none of the three ratings performed at that scale within the ten year period, thus it is deemed unnecessary to include it in the analysis. After

CEU eTD Collection

scoring each scale I summed up the values for each indicator separately. The results are shown in fig.6.

Figure 6. Bank , Borrower, Project Outcome performance converted into numerical values for comparison purposes



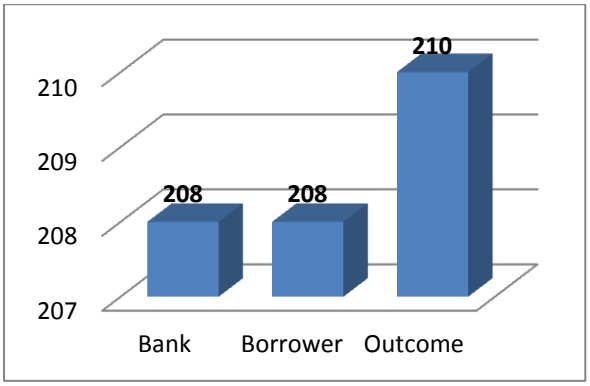
Source: Author's calculation based on tab.5 scaling

The results in fig 6. show both bank and borrower performance have the same values, 208 points. Whereas, the outcome performance was rated higher compared to them, with 234 points. It seems not possible to have the bank and the borrower performance rated lower, whereas the outcome performance higher. However, there are two key important facts to note here. First, outcome performance is not automatically translated to equal bank and borrower performance, rather according to OED (2005) it shows the extent to which the objectives were met efficiently. Thus, even if the project outcome is few points higher it is ok. The second major important issue is that some data are missing for the bank and borrower performance, whereas the outcome performance has values for each project. This means that more projects were included in outcome performance and less in bank and borrower performance, thus inflating the value of outcome performance in this case. The data was missing for seven projects, which when converted into percentages means that 10% of the projects did not have ratings for bank and borrower performance. In this case the only solution in order to compare the three ratings was to take out of the analysis the same seven projects from outcome performance that were missing in

CEU eTD Collection

bank and borrower performance. This way the number of included projects was the same in each rating. In this case, as fig.8 shows the difference is much smaller, it is only two points. Also this two point is possible considering the first fact stated above that says outcome does not automatically equal bank and bower performance according to the bank's definition.

Figure 7. All Bank, Borrower, Outcome performance converted into numerical values for comparison purposes



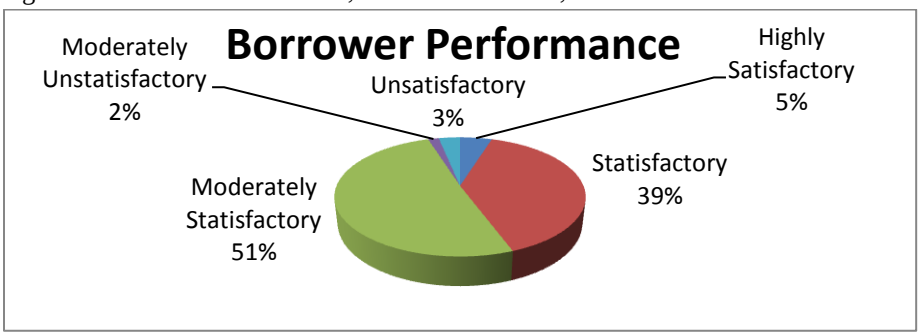
Source: Author's creation from author's database based on WB dataset

The actual results show that IEG rated the performance of project outcome, bank and borrower at the same level. According to these results, IEG was not biases toward the bank, but rated it equally with the borrower, and the project outcome. The predicted results based on literature fail to show that IEG as an internal body is less objective. On the contrary, IEG according to these data can be considered a complete independent body which also matches the findings in the Self-Evaluation report (2011) stating that IEG independency is in compliance with ECG criteria.

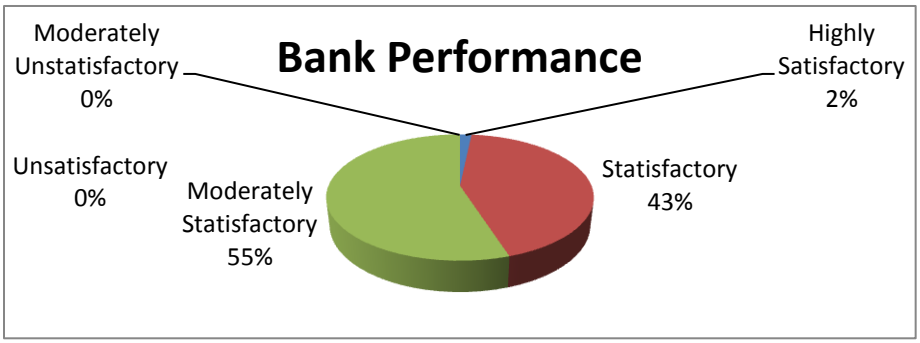
Furthermore, from the data calculations I could derive another output, the % of projects rated according to each scale, for the three rating including borrower (a), bank (b) and project outcome (c).

CEU eTD Collection

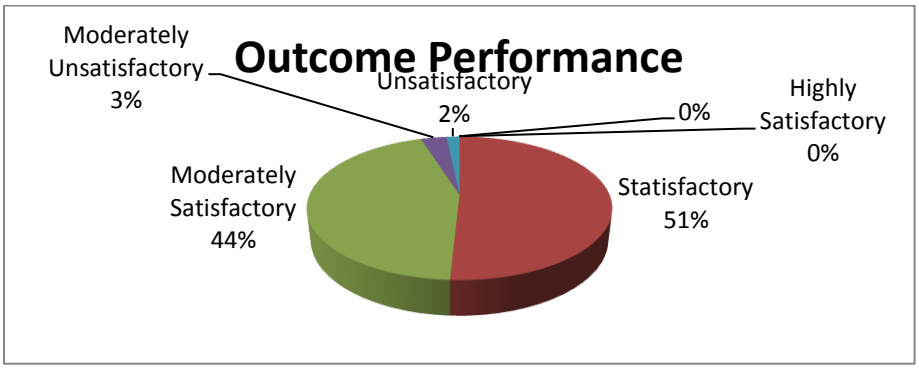
Figure 8. Borrower Performance, Bank Performance, and Outcome Performance



a.



b.



c.

Even from fig.8 a. b. c., it is evident that without assigning numerical values to each scale (as this paper did above), it is not possible to see from the % which rating performed better. For instance, misleading could be the results in fig. 8 a. and b. showing that 5% of projects rated to the highest scale the borrower performance, while only 2% from the bank performance. However, when all these are converted into numerical values as shown above, they perform exactly the same. This shows the importance of assigning numerical values for measuring and comparison, and it also shows the independency of the IEG.

CEU eTD Collection

- Spending per Sector Measured in $ and in %, and General Sector Performance

The second approach tends to derive the output in terms of sector performance and spending as total amount and as percentage. Each of the 68 closed evaluated projects is spread into several sectors. Sectors include a variety of areas including economic, law, trade, industry, telecommunication, mining, agriculture, energy, security, public administration, health, central administration, labor, social welfare, and the like. Running any analysis with all these categories would not show any specific results. Therefore, after getting the % spent on each sector from each project based on the WB database, I created my own categories which are fewer compared to WB sector categorizations. Trying to put similar sectors in the same category, and looking deeper into evaluation reports both, ISR and ISCR, it was evident that all projects collaborated with ministries. Therefore, it sounded rationale to create similar sector categories to types of ministries. Tab.6 shows the 10 types of categorizations I created for further analysis.

Table 6. Categorization of Projects into less sectors

Mining, Energy	Health, Labor, Social Welfare	Education	Economy, Finance, Banking	Public Administration	Justice, Law, Security	Agriculture, Forestry, Rural	Infrastructure, IT	Trade, Industry	Other

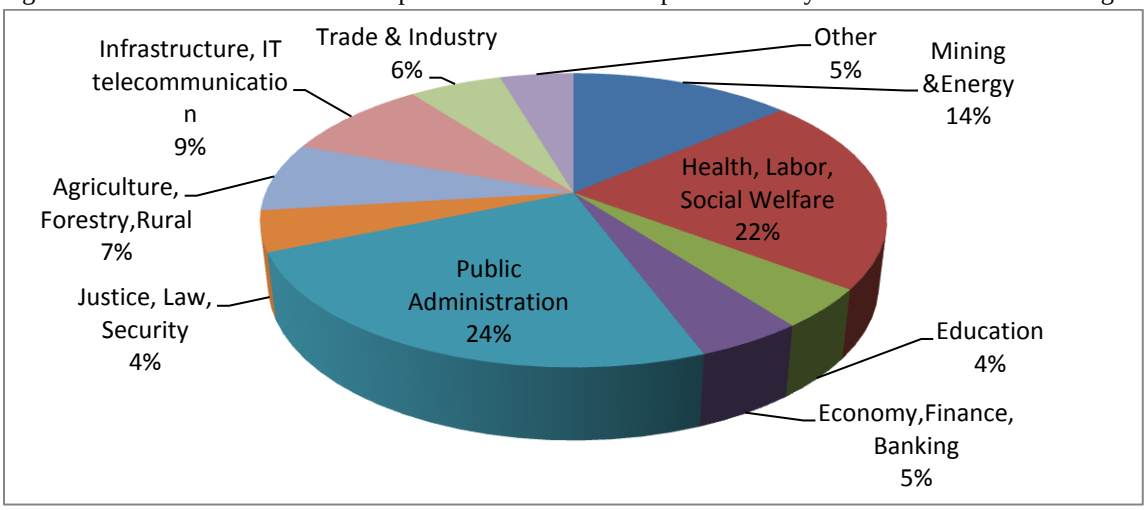
Source: Author's creation for comparison and analytical reasons

After categorizing all the data the first information to get from the data is the possibility to calculate which category expenditures were highest. Considering the goals and mission of WB to "Help Reduce Poverty" and "Support Development" (WB 2013), the predicted results argue that the highest investments are in three sectors Health, Labor, Social Welfare, and Education, and

CEU eTD Collection

Public Administration. The results in fig.9 show in percentage expenditures in each of the ten sectors.

Figure 9. % of commitment amount spent in each sector for a period of ten years in Western Balkan Region



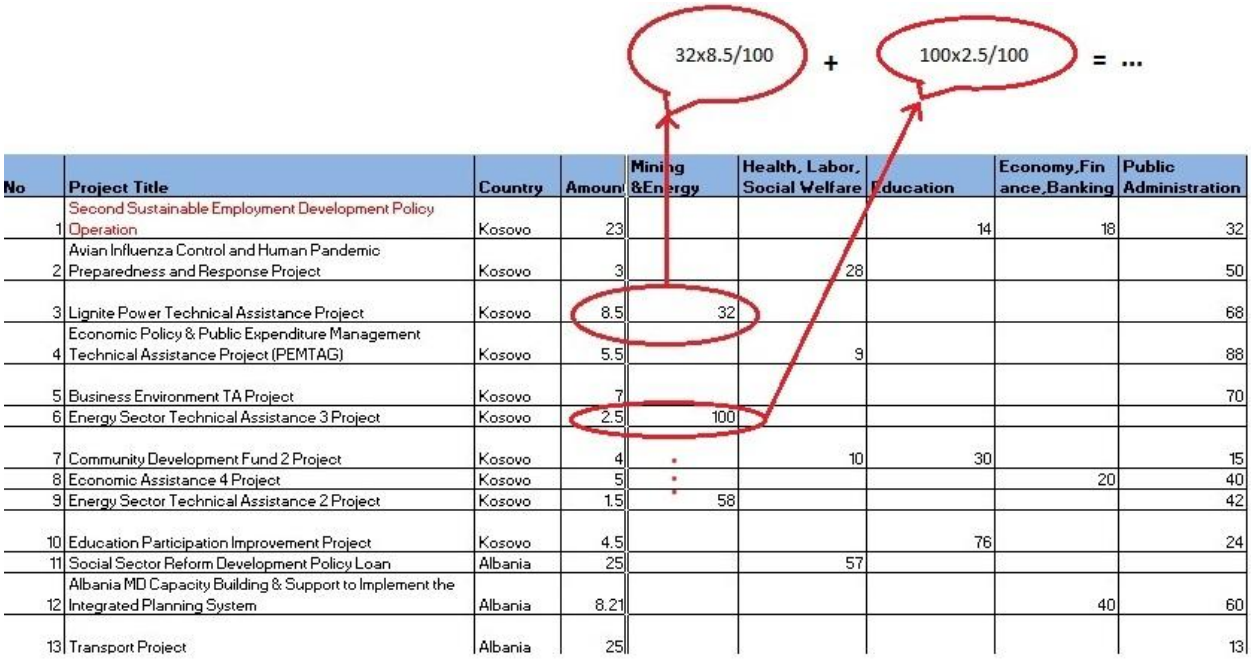
Source: Author's calculation from the database created based on the WB data-set

The results in fig.9 show that the highest investments took place in the sector of Public Administration, followed by Health, Labor, Social Welfare. However, the Education sector presents one of the lowest expenditures among other sectors together with Justice, Law, Security sector. The third highest sector in spending is the Mining & Energy sector which does not match the predicted results which were derived based on WB mission and objectives. Through mining and energy the bank cannot directly and easily facilitate development and reduce poverty as it is stating in its mission and objectives. Education is much more important in this regard. Does this mean that the attention of WB and objectives will shift in the coming years to other sectors? If yes, it would be interesting to know why and from when this shift? This question is out of the scope of this paper's analysis, but I think this would be a good research question for another paper.

CEU eTD Collection

Furthermore, it is important to calculate the amount spent in $, not only as % (as it was calculated above), in order to see if the rating sector investments will change. Amounts are added per each sector separately as shown below in fig. 10. Whereas the amount of expenditures is shown in tab.7.

Figure 10. The way calculations were performed to find out the amount spend in each sector



No	Project Title	Country	Amoun	Mining &Energy	Health, Labor, Social Welfare	Education	Economy,Finance,Banking	Public Administration
1	Second Sustainable Employment Development Policy Operation	Kosovo	23			14	18	32
2	Avian Influenza Control and Human Pandemic Preparedness and Response Project	Kosovo	3		28			50
3	Lignite Power Technical Assistance Project	Kosovo	8.5	32				68
4	Economic Policy & Public Expenditure Management Technical Assistance Project (PEMTAG)	Kosovo	5.5		9			88
5	Business Environment TA Project	Kosovo	7					70
6	Energy Sector Technical Assistance 3 Project	Kosovo	2.5	100				
7	Community Development Fund 2 Project	Kosovo	4		10	30		15
8	Economic Assistance 4 Project	Kosovo	5				20	40
9	Energy Sector Technical Assistance 2 Project	Kosovo	1.5	58				42
10	Education Participation Improvement Project	Kosovo	4.5			76		24
11	Social Sector Reform Development Policy Loan	Albania	25		57			
12	Albania MD Capacity Building & Support to Implement the Integrated Planning System	Albania	8.21				40	60
13	Transport Project	Albania	25					13

Source: Author's part of the database created based on WB data-set

Table 7. Spending in each sector, for all evaluated projects within a ten year period in Western Balkan Region

Mining, Energy	Health, Labor, Social Welfare	Education	Economy, Finance, Banking	Public Administration	Justice, Law, Security	Agriculture, Forestry, Rural	Infrastructure, IT	Trade, Industry	Other
163.62	303.82	18.70	179.72	248.35	62.16	55.88	139.73	133.77	42.99

Source: Author's calculation based on WB dataset; Note: all values are in million $

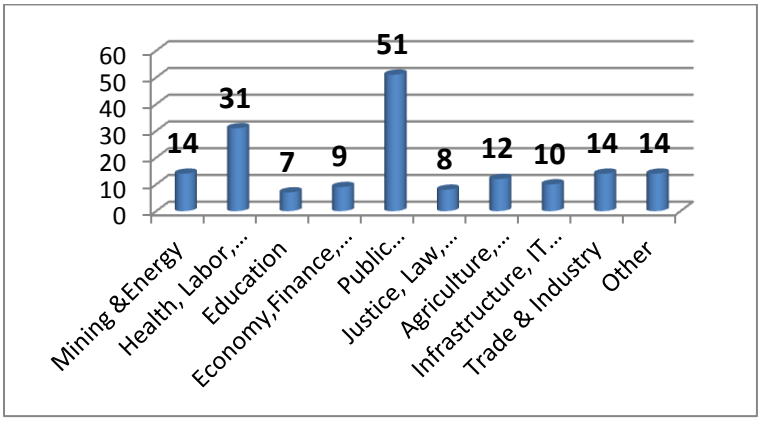
As shown in tab.7 Public Administration sector is not the sector with highest investments, but the first is the Health, Labor, Social Welfare, sector followed by Public Administration. This does

CEU eTD Collection

not represent a major change because they are both the two first sectors with highest investments. At the same time, the Education sector still remains the lowest sector invested in with only $18.70 mil. Also, the difference between Education and the other lowest spending sectors is a considerable amount of $24.29mil. The only change is Economy, Finance, Banking that moves up for few millions before the mining and energy sector, but this is not changing any results. Thus, the question about matching the mission of WB to reduce poverty and the shift investments in Mining &Energy sector, still remains.

Furthermore, from the data gathered this paper looks at another output in terms of sector performance. For finding out which sector performed better, the numerical values for each scale (see tab.5) are used, and are added up from every country on the same sector. Results are shown on the below fig.11.

Figure 11. Sector performance



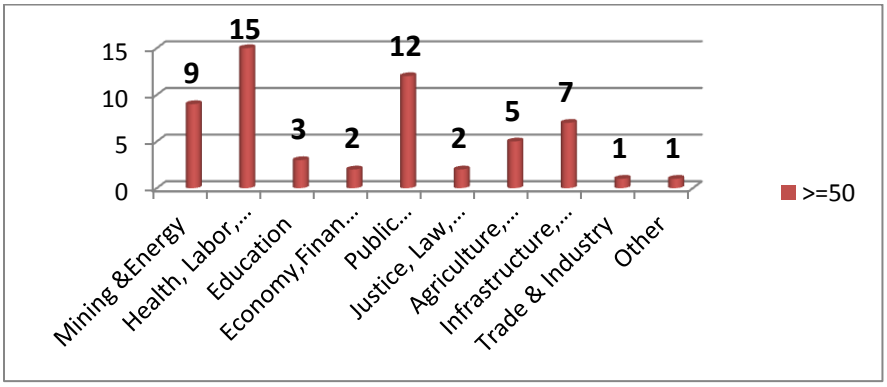
Source: Author's calculation from the database created based on WB data-set

The results in fig.11 show that Public Administration sector is the best performer, followed by Health, Labor, Social Welfare sector. Then, we have three other sectors Mining and Energy, Trade and Industry, and Other that are scored the same, 14 points. However, it is important to note that Public Administration and Health, Labor, Social Welfare are the sectors with the

CEU eTD Collection

highest percentage of investments followed by the Mining and Energy sector (see tab.7). The limitation of having different amounts of investments spend in each sector, does not clearly show if Public Administration is the best performer. In order to have more exact results by finding out the number of projects in each sector I have categorized each project in a sector. Since, the commitment amount for each project was spread into 3,4, or 5 other sectors, I have set a threshold of 50% or more. So each project that has spend 50% or more of its commitment amount in a sector was categorized in that sector. There is a limitation in here as well, since eleven projects could not be categorized, so they are left out of analysis. It was not possible to categorize these projects because of their uneven spread among many sectors, thereby I run the data-mining without these eleven projects. The results of these analysis are shown in fig.12.

Figure 12. Categorized projects



Source: Author's calculation

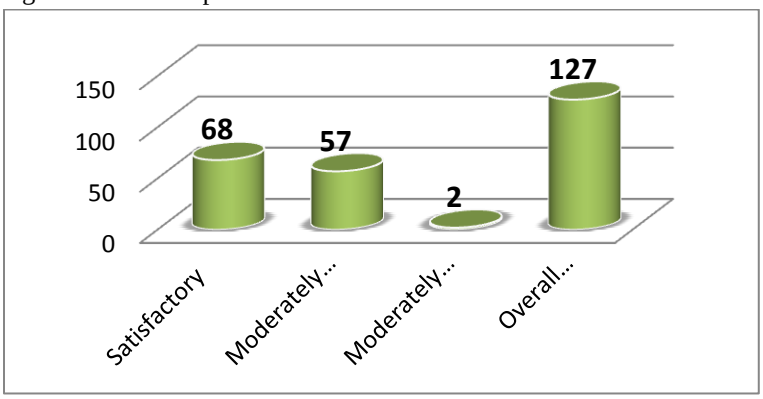
The results in fig.12 show Health, Labor, Social Welfare as the sector with the highest number of projects followed by Public Administration. These results back up the results shown in fig.11 and tab.7, claiming that these two sectors Public Administration and Health, Labor, Social Welfare are the first two most invested sectors in term of dollars, and as percentages of total amount invested.

CEU eTD Collection

- The Effect of Intermediate Results Indicator on Overall Project Outcome

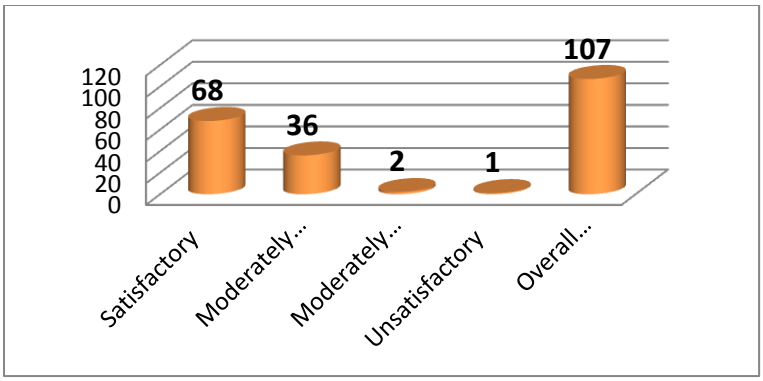
Third, this paper analyzes the effect of having or not having intermediate results indicator in outcome performance. These outputs can be derived due to data available in evaluation project reports. The first note is that not each evaluated project had intermediate results indicator. 31 out of 68 evaluated projects, did not have intermediate results indicator. The predicted results argue the projects that had intermediate results indicator also have positive effects on project outcome. This is based on the purpose of intermediate results indicator which are conducted in the mid-phase of project cycle to check if the project is going as planned, thus provide assistance if needed. With the purpose of finding that answer, the same numerical values were used with the exclusion of two scales highly satisfactory and highly unsatisfactory because none of the projects was rated with any of these two scales. Fig.13 a. shows the results when the intermediate results indicator were present, whereas b. shows the results when there was no intermediate results indicator.

Figure 13. Overall performance when intermediate results indicator were present (a) and when they were missing (b)



a.

CEU eTD Collection



b.

Source: Based on author's calculation

From the results shown in fig. 13 a) and b) the case when projects had intermediate indicator results the overall scoring seems to be higher by 20 points. Yet, this difference is not considerable when taking into account that under a. the data includes 37 projects whereas under b. only 31 projects. This difference in the number of projects that have intermediate results indicator inflates the overall points. Under current circumstances, nothing is possible to be done to make them more comparable. Therefore, this can be summed up by saying that it is not possible to see from the available data if these indicators have a positive impact on the overall project outcome.

- Country Performance

The fourth and last possible output to interpret from the data available in report is the country performance. According to IEG ratings country performance refers to the borrower performance. As discussed in chapter two (OED 2005), borrower performance refers to the extent to which it shows willingness and ability to guarantee and comply with the criteria and requirements agreed with the bank. This depends on other indicators such as corruption, political or economic

CEU eTD Collection

stability and the like. Therefore, in order to form the predicted results this paper looks at scoring of each Western Balkan country in the Corruption Perception Index (Transparency International 2012), the Doing Business (WB 2012), the Democracy Score (Freedom of House 2012), as shown in tab. 8.

Table 8. Country performance according to three indicators

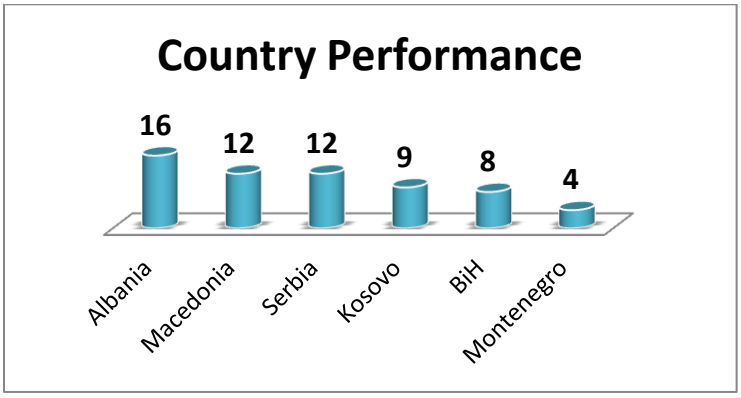
Nr	Country	Democracy Score Freedom of House 2012	Doing Business WB 2012	Corruption Perception Index Transparency International 2012
1	Albania	4.14	85	33
2	Bosnia and Herzegovina	4.36	126	42
3	Macedonia	3.89	23	43
4	Montenegro	3.82	51	41
5	Kosovo	5.18	98	34
6	Serbia	3.64	86	39

Source: Author's creation based on WB, Freedom of House, and Transparency International data
Note: WB- a higher ranking shows a more conducive environment to run a business; Freedom of House-1 represents the highest level of democracy; Transparency International- the higher the number the less corrupt is the country

From these available indicators in it not easy to say exactly which country is a better performer however we has BiH that leads in Doing Business, with a very high difference from other countries, and lacks only one point behind in Transparency International, whereas in Democracy score performs at a lower level compared to other countries. On the other hand, Serbia has the best rating for democracy index among other Western Balkan countries, but in two other indicators has a much lower scale as compared to BiH. Thus, the predicted results show that BiH, or Serbia might be better performer as opposed to others. In order to find the answer this paper uses the same scales with same numerical values, and the same sector categorizations, to calculate country performance. The results are shown in fig. 14.

CEU eTD Collection

Figure 14. Country performance



Source: Author's calculation based on the data from WB

The results in fig.14 show that Albania is the best performer, followed by Macedonia and Serbia. These results do not match in any way the predicted results. However it is important to also note the number of projects, in order to compare it with performances. Going back to tab.4 we can see that number of evaluated projects in Albania is also the highest, followed by Serbia. This does not exactly mean that Albania had the best performance followed by Serbia, because they both were also leading with the number of projects, thus more projects adds more points in the calculations in favor of these countries. To sum up, it is clear which country performs best according the data available in evaluation reports. Following these discussions about different outputs derived and interpreted from the reports, the next section will summarize the findings of this chapter.

3.4 Summary of Findings

The first category of findings, show through data gathering that not all projects implemented within the ten year period in the bank were evaluated. Therefore, before analyzing the evaluated projects, this paper tried to identify whether there is any visible criterion that decides which

CEU eTD Collection

project to be evaluated or that makes a project have higher chances of being evaluated because of possessing a particular criterion. Findings in this category show that the commitment amount does not influence neither negatively, nor positively, the likelihood of a project to be or not to be evaluated. But, the results show that there is a higher likelihood to evaluate a project if that particular project has intermediate results indicator.

Second group of findings included only the evaluated projects. First, output to be interpreted by looking at IEG ratings of the bank, the borrower and the overall project performance in the reports is the high degree of IEG independency. This degree of independency also proves accountability to the donor, because this shows that IEG was subjective when it evaluated the projects. Second, the findings show that having intermediate results indicator does not influence neither positively nor negatively the overall project performance. Third, although not so visible but after certain calculation the Mining and Energy sector came out as the third highest in terms of the amount of money spend by the bank in these kind of projects. This finding, did not match the predicted results based on the bank objective. The fourth output assessed was country performance. At first Albania was the best performer and this results did not match the predicted results based on well-known indicators such as WB Doing Business. However, when looking at the number of projects in each country, Albania had also the highest number of projects, thus the high points in performance were due to high number of projects, as well. This shows that from the available data it was not possible to identify which country is the best performer. To sum up, these reports are informative to a high degree because as presented above there is quite a lot of output that can be interpreted after some data gathering and calculations. These findings apply particularly to Western Balkan region, but taking into consideration that this region presents all

CEU eTD Collection

types of projects that the bank implements worldwide, there is a high likelihood that the same results would be found in other regions, as well.

CEU eTD Collection

Implications of the Research

The importance of evaluating developmental activities has increased over time. The main objective of this paper was to assess how informative are the WB, IEG project evaluations in the sense of output that can be interpreted from them. To reach this aim, this paper looked at implemented projects in Western Balkan, for a period of ten years, because Western Balkan represents all types of projects that the bank implements worldwide making this region an interesting and comprehensive subset to analyze. Project evaluations of WB were chosen to be assessed because the bank is one of the leading organizations in evaluation and one of the largest multilateral organizations. This paper created its own data-set by gathering the information needed to answer the research question from the WB database. It has also run data-mining while giving numerical values to performance scales in order to compare them. Additionally, less sector categorizations were formed with the purpose of categorizing each project in a sector. A threshold was also set, so that each project that invested 50% or more in one sector, was categorized in that particular sector. The data analysis were approached quantitatively to interpret certain outputs through the use of EXCEL and Microsoft SQL Server 2008 express edition. These two programs made the data more readable and understandable.

The data gathering has shown that not all implemented projects were evaluated and there was no evidence showing why some projects were evaluated and some others not. Thus, before analyzing the evaluated projects this paper found it necessary to identify if possible whether there is any criteria that makes a project more likely to be evaluated as opposed to another. The

CEU eTD Collection

analyses looked at two variables, commitment amount and intermediate results indicators. The results show that evaluated projects range within different commitment amounts establishing no relation between that and the decision to evaluate or not a project. But, another result shows that there is a higher likelihood for a project to get evaluated if that project has also the intermediate results indicators.

The second group of results show that evaluation reports of developmental projects in the bank were quite informative in terms of showing how well the bank, the borrower, and the project outcomes performed. This also suggested that the independency of the IEG is very high. Also, results show that no relation can be established between intermediate results indicator and overall outcome project performance. Another output that could be derived from evaluations after certain calculations were sector expenditures and performances. The results showed that Public Administration and Health, Labor, Social Welfare, represent the highest amount of expenditures and are best performers followed by the Mining and Energy sector. To analyze why the Mining and Energy sector came out as the third highest sector but which does not match the bank's objectives is out of the scope of this analysis. Analyzing this investment shift could be a good research question for another paper. The last output looked at was country performance. Results showed that Albania was the best performer followed by Serbia, yet the number of projects was also highest in these two countries. However, the country performance could not be traced at earlier stages to find out the rationale for having Albania as the best performer. But, analyzing why some countries have higher number of projects compared to others would be an interesting research question, and would probably clarify the puzzle of Albania being the best performer. In conclusion, these findings apply particularly to Western Balkan region, but taking

CEU eTD Collection

into consideration that this region presents a comprehensive range of projects that the bank implements, there is a high likelihood that the same results would be found in other regions, as well.

CEU eTD Collection

Bibliography

Asian Development Bank (2006). Impact Evaluation: Methodological and Operational Issues. *Published in Philiphines.*

Barton, T. (1997). Guidelines to Monitoring and Evaluation. How are we doing? *CARE International, Uganda.*

Burnside, A.C. and D. Dollar (2000). Aid, policies, and growth. *American Economic Review.*

Baker, J. (2000). Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners. *Washington: World Bank.*

Bamberger, M. (2000). The Evaluation of International Development Programs: A View from the Front. *American Journal of Evaluation.*

Bamberger, M. (2009). Institutionalizing Impact Evaluation Within the Framework of a Monitoring and Evaluation System. Poverty Analysis, Monitoring, and Impact Evaluation Thematic Group. *The World Bank.*

Collier, P. and D. Dollar (1998). Aid Allocation and Poverty Reduction. *Washington DC: World Bank*

Cracknell, B. (2000). Evaluating Development Aid: Issues, Problems, and Solutions. *Sage Publications: New Delhi.*

Crawford, P., Bryce,P. (2003). Project monitoring and evaluation: a method for enhancing the efficiency and effectiveness of aid project implementation. *International Journal of Project Management.*

Checkland, P. B. (2001). Soft Systems Methodology, in J. Rosenhead and J. Mingers (eds) *Rational Analysis for a Problematic World Revisited. Brisbane: Wiley*

Crawford, P., Perryman, J., Petocz, P.(2004). Synthetic Indices: A Methods for Evaluating Aid Project Effectiveness. *Sage Publication.*

Chelimsky, E. (1987). The Politics of Doing Program Evaluation. *Society, Springer, New York.*

Chianca, T. (2008). The OECD/DAC Criteria for International Development Evaluations: An Assessment and Ideas for Improvement. *Journal of Multidisciplinary Evaluation.*

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. (2007). Are Aid Agencies Improving? *Economic Policy.*

CEU eTD Collection

Freedom of House (2012).Rating and Democracy Score, Nations in Transit. http://www.freedomhouse.org/sites/default/files/2012%20%20NIT%20Tables.pdf

Grun, R.(2006). Monitoring and Evaluating Projects: A step by step Primer on Monitoring, Benchmarking, and Impact Evaluation. *Health, Nutrition, and Population (HNP), World Bank publishing.*

Grasso, P., Wasty, S., Weaving, R. (2003). World Bank Operations Evaluation Department. The First 30 Years. *WB publishing, Washington DC.*

Hailey, J., James, R. Wrigley, R. (2005). Rising to the Challenges: Assessing the Impacts of Organizational Capacity Building. *Praxis Paper No.2 : The International NGO Training and Research Center.*

Independent Evaluation Group (2011). Results and Performance of the World Bank Group. http://ieg.worldbankgroup.org/content/ieg/en/home/reports/rap2011.html

Independent Evaluation Group (2013). Improving World Bank Group Development Results through Excellence in Evaluation. About IEG. http://ieg.worldbankgroup.org/content/ieg/en/home/about.html

Independent Evaluation Group (2011). Access to Information Policy. IEG: WB,IFC, MIGA publishing.

Independent Evaluation Group. Access to Information. http://ieg.worldbankgroup.org/content/ieg/en/home/access_to_information.html

Independent Evaluation Group (2013). Evaluation Week. http://ieg.worldbankgroup.org/content/ieg/en/home/Events/evaluationweek.html

Independent Evaluation Group (2011). Self-Evaluation of the Independent Evaluation Group. World Bank Publishing, Washington, D.C.

Independent Evaluation Group (2013). World Bank Project Performance Ratings. http://ieg.worldbankgroup.org/content/ieg/en/home/ratings.html

Jeffrey, S. (2005). The End of Poverty: Economic Possibilities for Our Time. *New York: Penguin Press.*

Labs, E. (1997). The Role of Foreign Aid in Development. *The University of Michigan.*

Moyo, D. (2009). Dead Aid. Why Aid is Not Working and How There is Another Way for Africa. *Penguin Books.*

Mohr, L.B. (1995). Impact Analysis for Program Evaluation. *2nd ed. Thousand Oaks, Calif.: Sage Publications.*

CEU eTD Collection

Marinaccio, M. and Trojanowski, M. (2012). Projects, Programs Defined. *Internal Auditor. Illustration by Sean Yates/ Yacinsky Design LLC.*

Naonobu, M. and Nobuko, F. (2009). Evaluating Development Assistance: A Japanese Perspective. *Foundation for Advances Studies on International Development, International Development Research Institute, Tokyo, Japan.*

OECD (2010). Evaluation in Development Agencies, Better Aid. *OECD Publishing.*

OECD, DAC Network on Development Evaluation (2010). Quality Standards for Development Evaluation. *OECD Publishing.*

OECD (1991). Principles for Evaluation of Development Assistance. *Development Assistance Committee, Paris.*

OECD (2010). Quality Standards for Development Evaluation. DAC Guidelines and References Series. *OECD Publishing.*

Operations Evaluation Department (2005). Harmonized Evaluation Criteria for ICR and OED Evaluations. *WB Publishing, Washington, D.C.*

Pacquement, F. (2010). How Development Assistance from France and the United Kingdom Has Evovled: Fifty Years from Decolonisation. *Geneva- Graduate Institute of International and Development Studies.*

Palumbo, D., Nachmias, D. (1983). The Preconditions for Successful Evaluation: Is There an Ideal Paradigm? *Policy Science: Elsevier Science Publisher B.V., Amesterdam.*

Project Implementation and Status Results (PISR) (2012). http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/ECA/2012/06/27/C1413BC47957F75985257A2A006FBB10/1_0/Rendered/PDF/ISR0Disclosabl02720120134082842 3956.pdf

Pitman, G., Feinstein, O., Ingram, G. (2005). Evaluating Development Effectiveness. World Bank Series on Evaluation and Development. *Volume 7, Transaction Publishers. New Brunswick and London.*

Riddell, R. (2007). Does Foreign Aid Really Work? *Oxford University Press.*

Scriven, M. (2007). The Logic of Evaluations. *Department of Psychology, Claremont Graduate University. USA.*

Scriven, M. (1967). The Methodology of Evaluation. In Perspectives of curriculum evaluation, eds. Ralph W. Tyler, Robert M. George. *AERA Monograph Series on Curriculum Evaluation, vol. 1. Chicago: Rand McNally.*

CEU eTD Collection

Skoufias, E., Parker, S. (2001). Conditional Cash Transfers and Their Impact on Child. *Discussion Paper, International Food Policy Research Institute, Washington, D.C.*

Taylor, J. (2003) Using Measurement Developmentally. *Cape Town: CDRA.*

Tuan, M. (2012). External Evaluation Advisory Committee Scoping Project: Findings and Recommendations. *The David and Lucile Packard Foundation.*

TRANSPARENCY INTERNATIONAL (2012). Corruption PERCEPTION INDEX http://cpi.transparency.org/cpi2012/results/

The Center for Global Development (2006). When Will We Ever Learn? Improving Lives through Impact Evaluation. *Evaluation Gap Working Group. Washington, D.C.*

UNDP (1997). Results-oriented Monitoring and Evaluation. *UNDP publishing.*

USAID (2000). Measuring Institutional Capacity. Recent Practices in Monitoring and Evaluation Tips, Number 15. *Washington, DC. Center for Development Information and Evaluation.*

Weiss, C. (1998). Evaluation. 2nd ed. Prentice Hall, Upper Sadler River, New Yersey.

Weiss, C. (1973). Where Politics and Research Evaluation Meet. *American Journal of Evaluation.*

World Bank, IPDET, Presentation. http://www.worldbank.org/oed/ipdet/presentation/M_01-Pr.pdf

World Bank. Archives. http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/EXTARCHIVES/0,,contentMDK:23164816~pagePK:36726~piPK:36092~theSitePK:29506,00.html

Work and Schooling: Evidence from the PROGRESA Program in Mexico. *Project Muse. Published by Brookings Institute Press.*

World Bank (2013). Home, Who we are, Mission. http://web.worldbank.org/WBSITE/EXTERNAL/EXTABOUTUS/0,,contentMDK:20046292~menuPK:1696892~pagePK:51123644~piPK:329829~theSitePK:29708,00.html

World Bank (2012). Doing Business Report. http://www.doingbusiness.org/rankings

World Bank (2004). Monitoring & Evaluation: Some Tools, Methods & Approaches. *World Bank Operations Evaluation Department, Evaluation Capacity Development. Washington DC.*

World Bank (2006). Doing Impact Evaluation: Impact Evaluation and the Project Cycle. *Word Bank Publishing.*

CEU eTD Collection

World Bank (2013). Projects and Operations Dataset. http://www.worldbank.org/projects

CEU eTD Collection