This report synthesized the completed evaluation studies on educational policy conducted by the World Bank from 1990-2006. There were 28 studies identified within the area of evaluation of educational policies. Through vote counting the studies were synthesized according to the following classifications: (1) Number of evaluation studies completed each year; (2) areas of investigation; (3) evaluation study for each country; (4) methods used for each evaluation study; and (5) models of evaluation used. Based on the findings, four recommendations were provided for future educational evaluation studies.

Keywords: Evaluation studies, educational policies, evaluation models

World Bank Evaluation Studies on Educational Policy

Carlo Magno
De La Salle University, Manila, Philippines

This report provides a panoramic view of different studies on education sponsored by the World Bank focusing on the evaluation component. The report specifically presents completed studies on educational policy from 1990 to 2006. A panoramic view of the studies was presented showing the area of investigation, evaluation model, method used, and recommendations. A synthesis of these reports was shown in terms of the areas of investigation, content, methodology, and model used through vote counting. The vote counting is a modal categorization assumed to give the best estimate of selected criteria (Bushman, 1997).

The World Bank provides support to education systems throughout the developing world. Such support is broadly aimed at helping countries attain the objectives of "Education for All" and education for success in the knowledge economy. An important goal is to tailor bank assistance to region- and country-specific factors such as demographics, culture, and the socio-economic or geopolitical climate. Consequently, the top priority is to inform development assistance with the benefit of country-specific analysis

examining (1) what factors drive education outcomes; (2) how do they interact with each other; (3) which factors carry the most weight and which actions are likely to produce the greatest result; and (4) where do the greatest risks and constraints lie. The World Bank classify the countries according to different regions such as Sub-Saharan Africa, East Asia and the Pacific, Europe and Central Asia, Latin America and the Carribean, and Middle East and North Africa.

Areas of Investigation

There are 28 studies done on educational policy with an evaluation component. Education studies with no evaluation aspect were not included. The different areas of investigation were enumerated and the number of studies conducted for each country according to the sequence of years was counted (see Table 1). Most of the studies on educational policy are targeting the basic needs of a country and specified region of the world such as the effectiveness of education in the basic education, tertiary, critical periods such as child development programs, and promoting adult literacy. From the earliest period (1990's) the trend of the studies done are on Information and Communications Technology (ICT) on basic education. The pattern for the 21st century studies shows a concentration in evaluating the implementation of tertiary education across countries. This is critical since developing nations rely on the expertise produced by its manpower in the field of science and technology. For the latest period, a new area of investigation which is language learning was explored due to the recognition of globalization on some countries like Vanuatu.

It is shown in Table 1 that most studies on educational policy were conducted for the year 2000 since it is a turning point of the century. For the coming of a new century much is being prepared and this is operationalized by assessing a world wide report on what has been accomplished from the recent 20th century. The studies typically cover a broad range of education topics such as school self-evaluation, early child development, basic education, adult literacy, and tertiary distance education. These areas of investigation cover most of the fields done for the 20th century and an overall view of what has been accomplished was reported. It can also be noted that there is an increase of studies conducted at the start of the 21st century. This can be explained with the growing trend in globalization where communication across countries is more accessible. It can also be noted that no studies were completed on educational policy with evaluation for the years 1993, 1995, 1997, and 2005. The trend in the number of studies shows that consequently, after a year, a study gives more generalized findings since it covered a larger and wide array of sampling where a long period of time was taken to finish. More results are expected before the end of 2005. The trend of studies across the years is significantly different with the expected number of studies as revealed using a one-way chi-square where the computed value (χ^2=28.73, *df*=14) exceeds a

probability of error. There is a significant large concentration of evaluation studies conducted in the year 2000.

Table 1
Counts of Area of Investigation From 1990 - 2006

Year	Country	Area of Investigation	No. of Studies	Total no. of Studies per year
2006	Vanuatu	Language learning	1	1
2005		None	0	0
2004	Indonesia	Undergraduate/Tertiary Education	2	5
	Thailand			
	Senegal	Adult Literacy	1	
	Different Regions, Columbia	Early Child Development	2	
2003	Thailand	Undergraduate/Tertiary Education	1	2
	Different Regions	AIDS/HIV Prevention	1	
2002	Different regions	Textbook/Reading materials	1	2
	Africa	Secondary Education	1	
2001	Brazil	Early Child Development	1	2
	China	Secondary Education	1	
2000	Different Regions	School Self-evaluation	1	7
	Different Regions	Early Child Development	1	
	Pakistan, Cuba	Basic Education	3	
	Africa	Adult Literacy	1	
	Africa	Tertiary Distance Education	1	
1999	USA	Test Evaluation	1	3
	Different Regions	Infant Care	1	
		Early Child Development	1	
1998	Different Regions	Teacher Development	1	2
	Different Regions	ICT	1	
1997		None	0	0
1996	Different Regions	Basic Education (school financing)	1	2
	Chile	ICT	1	
1995		None		0
1994	Philippines	Vocational Education	1	1
1993		None		0
1992	Different Regions	Secondary Education	1	1
				Total=28

Table 2
Counts of Area of Investigation From 1990 - 2006

Area of Investigation	Number of Studies
Language learning	1
Undergraduate/Tertiary Education	4
Adult literary	2
Early Child Development	5
AIDS/HIV Prevention	1
Textbook/Reading material	1
Secondary education	3
School Self-evaluation	1
Basic education	4
Test Evaluation	1
Infant Care	1
ICT	2
Teacher Development	1
Vocational Education	1

Table 2 shows the number of studies conducted for every area in line with educational policy with evaluation. Most of the studies completed and funded are in the area of early child development followed by tertiary education and basic education. This can be explained by the increasing number of early child care programs around the world that needs to be evaluated in terms of its effectiveness at a certain period of time. Much of the concern is on early child development since it is a critical stage in life which evidently results to hampering the development of an individual if not cared for at an early age. This also shows the increasing number of children where their needs are undermined and intervention has to take place. These programs sought the assistance of the World Bank because they need further funding for the program to exist. Having an evaluation of the child program likely supported the approval for further grant.

Somehow there are a large number of studies on basic and tertiary education where its effectiveness is evaluated. Almost all countries offer the same structure of education world wide in terms of the level from basic education to tertiary education. These deeply need attention since it is a basic key to developing nations which is to improve the quality of their education. The quality of people with skills depend on the countries overall labor force.

When the observed counts of studies for each area of interest is tested for significant differences, the computed chi-square value (χ^2=13, *df*=13) did not reach significance at 5% level. This means that the observed counts per area do not significantly differ to what is expected to be produced. There is at least an equal distribution of studies conducted within each area of investigation.

Table 3
Study Grants by Country

Country	No. of studies
Vanuatu	1
Indonesia	1
Thailand	1
Senegal	1
Different Regions	10
Brazil	1
China	1
Pakistan	1
Cuba	1
Africa	2
USA	1
Chile	1
Philippines	1

The studies done for each country are almost equally distributed except for Africa with two studies from 1990 until the present period. There is a bulk of studies done worldwide which covers a wider array of sampling across different countries. The world wide studies usually evaluate common programs across different countries such as teacher effectiveness and child development programs. There is great difficulty to come up with an efficient judgment of the overall standards of each program separately. The advantage of having a world wide study on educational programs for different regions means having a simultaneous description of the common programs that are running and the funding is most likely concentrated to one team of investigators rather than separate studies with different fund allocations. Another is the efficiency of maintaining consistency of procedures across different settings. The same standards of investigation were maintained for each country.

In the case of Africa, two studies were granted concentrating on adult literacy and distance education because these educational programs are critical in their country as compared to others. As shown in the demographics of the African region that their programs (adult literacy, distance education) are increasingly gaining benefits for its stakeholders. There is a report of remarkable improvement on their adult education and more tertiary students are benefiting from the distance education. Since the programs are improving, much funding is needed to continue the development.

When the number of studies are tested for significance across countries, the chi-square computed (χ^2=35.44, *df*=12) reached significance against a critical value of χ^2=21.03 at 5% probability levels. This means that the number of studies for each country differs significantly to what is expected to be produced. This is also due to having a large concentration of studies for different regions as compared to minimal studies for each country which made the difference.

Method of Studies

Various methodologies were used to investigate the effectiveness of educational programs across different countries. It can be seen in the reports that there is not much concentration and elaboration on the use and implementation of the procedures done to evaluate the programs. Most reports only mentioned the questionnaires and assessment techniques used. There are some that mentioned a broad range of methodologies such as quasi-experiments and case studies but the specific designs are not indicated. It can also be noted that the reports written by researchers/professors from universities are very clear in their method which is academic in nature but World Bank personnel writing the report tends to focus on the justification of the funding rather than the clarity of the research procedures undertaken. It can also be noted that the reports did not show any parts of the methodology. Most presented the introduction and some justifications of the program and later in the end the recommendations. The methodologies are just mentioned and not elaborated within the report and only mentioned on some parts of the justification of the program.

Table 4
Counts of Methods Used

Method	Counts
Questionnaires/Inventories/Tests	4
Quasi Experimental	5
True Experimental	1
Archival Data (Analyzed available demographics)	6
Observations	1
Case Studies	1
Surveys	1
Multimethod	9

It can be noted in Table 4 that most studies employ a multimethod approach. The multimethod approach creates an efficient way of cross-validating results for every methodology undertaken. One result in one method can be in reference to another result to another method which makes it powerful than using a single method. Since evaluation of the program is being done in most studies, it is indeed better to consider using a multimethod approach since it can generate findings where the researcher can arrive with better judgment and description of the program.

It can also be noted that most studies are also using archival data to make justifications of the program. Most of the researchers make reference to the archival data in coming up with inferences from enrollment percentage, drop out rates, achievement levels, and statistics on physical conditions such as weight and height etc. which can be valid. But they do not directly assess the

effectiveness of the program. The difficulty of using these statistics is that they do not provide a post measurement of the program evaluated. These may be due to the difficulty of arriving with national surveys on achievement levels and enrollment profiles of different educational institutions which is done annually but may not be in concordance with the timetable of the researchers. It is also commendable that a number of studies are considering the use of quasi-experimental designs to directly assess the effectiveness of educational programs.

The counts of the methodologies used is tested for significance, the computed chi-square value (χ^2=18.29, *df*=7) reached significance over the critical chi-square value of χ^2=14.07 with 5% probability of error. This shows that the methodologies used significantly varies to what is expected with high frequency on multimethod approaches.

The Use of Evaluation Models

The evaluation methods used by the studies were counted. There was difficulty in identifying the models used since the researchers did not specifically elaborate the evaluation or framework that they used. It can also be noted that the researchers are not really after the model but in establishing the program or continuity of the program. There are marked difference between university academics and World Bank personnel in doing the study. The personnel are not every elaborate in their assessment due to the lack of guidance from a model and academics would specifically state the context but somehow failed to elaborate in the process like adopting a CIPP model for example. Most studies are clear in their program objectives but failed to provide accurate measures of the program directly. The worst is that most studies are actually not guided with the use of a model in evaluating the educational programs proposed.

Table 5
Counts Models/Frameworks Used

Model/Framework	Counts
Objectives-Oriented Evaluation	10
Management-Oriented Evaluation	9
Consumer-Oriented Evaluation	0
Expertise-Oriented Evaluation	7
Participant-Oriented Evaluation	1
No model specified	3

As shown in Table 5 that majority of the evaluation used the objectives-oriented where they specify the program objectives and evaluated accordingly. A large number also used the management oriented and specifically made use of the CIPP by Stufflebeam (1968). A number of studies also used experts as external evaluators of the program implementation. Most of the studies

actually did not mention the model used and the models were just identified as described by the procedure in conducting the evaluation.

Most studies used the objectives oriented since the thrust is on educational policy and most educational programs start programs with clear objectives. These objectives are also treated as ends where the evaluation is basically used as the basis for the attainment. The other studies which used the management-oriented evaluation are the ones that typically describe the context of the educational setting where they used available archival data provided by national and countrywide surveys. The inputs and outputs are also described but most are weak in elaborating the process undertaken. The counts on the use of evaluation models (χ^2=18, *df*=5) reached significance at 5% level. This means that the counts are significantly different with the expected. This shows a need to use other models of evaluation as appropriate to the study being conducted.

Recommendations

1. It is recommended to increase distribution of study grants across countries. There is concentration of performing studies regionally which may neglect cultural and ethical considerations on testing and other forms of assessment. As a consequence there is no cross-cultural perspective on how the programs are implemented for each country because the focus is on the consistency of the programs. Conducting individual studies will show a more in-depth perspective of the program and how it is situated within a specific context.
2. It is recommended to have a specific section on the methodology undertaken by the researchers. This helps future researchers to qualify for the validity of the procedures undertaken by the evaluation study. Specifying clearly the method used enables the study to be replicated as best practices for future researchers and can easily identify procedures that needs to be improved.
3. It is recommended to have separate studies concentrating exclusively on program evaluation after successive program implementations. This will provide a better picture on the worth of a program. The judgment on how the program took place is focused which do not undermine the result of the program effectiveness. A good alternative is for the research grantee to allocate another budget on a follow up program evaluation after establishing the program.
4. It is recommended that when screening for studies, criteria on the use of an evaluation model should be included. The researchers making an evaluation study can be guided better with the use of an evaluation model.

References

Bray, M. (1996). *Decentralization of education community financing*. Washington, DC: World Bank Reports.

Brazil Early Child Development: A Focus on the Impact of Preschools. (2001). World Bank Reports.

Bregman, J., & Stallmeister, S. (2002). *Secondary education in Africa: Strategies for renewal*. Washington, DC: World Bank Reports.

Bushman, B. J. (1997). Vote-counting procedures in meta-analysis. In H. Cooper and Hedges, L. V. (Eds.), *The handbook of research synthesis*. New York: Russell Sage Publications.

Craig, H. J., Kraft, R. J., & du Plessis, J. (1998). *Teacher development: Making aniImpact*. Washington, DC: World Bank Reports.

Education and HIV/AIDS: A Sourcebook of HIV/AIDS Prevention Programs. (2003). Washington, DC: World Bank Reports.

Fretwell, D. I., & Colombano, J. E. (2000). *Adult continuing education: An integral part of lifelong learning emerging policies and programs for the 21st century in upper and middle income countries*. Washington, DC: World Bank Reports.

Gasperini, L. (2000). *The Cuban education system: Lessons and Dilemmas*. Washington, DC: World Bank Reports.

Getting an Early Start on Early Child Development. (2004). World Bank Reports.

Grigorenko, E. L., & Sternberg, R. J. (1999). Assessing cognitive Development In Early Childhood. Washington, DC: World Bank Reports.

Indonesia - Quality of Undergraduate Education Project. (2004). Washington, DC: World Bank Reports.

Liang, X. (2001).*China: Challenges of secondary education*. Washington, DC: World Bank Reports.

Nordtveit, B. J. (2004). *Managing public-private partnership lessons from literacy education in Senegal*. Washington, DC: World Bank Reports.

O'Gara, C., Lusk, D., Canahuati, J., Yablick, G., & Huffman, S. L. (1999). *Good practices in infant and toddler group care*. Washington, DC: World Bank Reports.

Operational Guidelines for textbooks and reading materials. (2002). Washington, DC: World Bank Reports.

Orazem, P. F. (2000). *The urban and rural fellowship school experiments in Pakistan: Design, evaluation, and sustainability*. Washington, DC: World Bank Reports.

Osin, L. (1998). *Computers in education in developing countries: Why and how?* Washington, DC: World Bank Reports.

Philippines - Vocational Training Project. (1994). Washington, DC: World Bank Reports.

Potashnik, M. (1996). Chile's Learning Network. Washington, DC: World Bank Reports.

Riley, K. & MacBeath, J. (2000). *Putting school self-evaluation in place*. Washington, DC: World Bank Reports.

Saint, W (2000). *Tertiary Distance Education and Technology and sub Saharan Africa*. Washington, DC: World Bank Reports.

Saunders, L. (2000). *Effective schooling in rural Africa report 2: Key issues concerning school effectiveness and improvement*. Washington, DC: World Bank Reports.

Stufflebeam, D. L. (1968). *Evaluation as enlightenment for decision making*. Columbus: Ohio State University Evaluation Center.

Tertiary Education in Colombia Paving the Way for Reform. (2003). Washington, DC: World Bank Reports.

Thailand - Universities Science and Engineering Education Project. (2004). Washington, DC: World Bank Reports.

Vanuatu: Learning and Innovation Credit for a Second Education Project. (2006). Washington, DC: World Bank Reports.

Ware, S. A. (1992). *Secondary school science in developing countries status and issues*. Washington, DC: World Bank Reports.

Xie, O., & Young, M. E. (1999). *Integrated child development in rural China*. Washington, DC: World Bank Reports.

Young, E. M. (2000). *From early child development to human development: Investing in our children's future*. Washington, DC: World Bank Reports.

Dr. Carlo Magno is a faculty of the Counseling and Educational Psychology Department, College of Education at De La Salle University, Manila. He specialize research in measurement and evaluation. Further correspondence can be sent to him at the same university 2401 Taft Ave. Manila, Philippines, e-mail: carlo.magno@dlsu.edu.ph