

PROGRAM THEORY DRIVEN EVALUATION OF THE RATIONALIZED UPGRADING OF THE DEPARTMENT OF SCIENCE AND TECHNOLOGY REGIONAL STANDARDS AND TESTING LABORATORIES IN REGION-X

Romela N. Ratilla¹

Abstract

This paper is a program theory driven evaluation of the project “Rationalized Upgrading of the DOST Regional Standards and Testing Laboratories” (RURL) conducted in December 2010 after 3.5 years of implementation. The project was implemented in the 14 regional offices with a total cost of P156M of which P9M was spent for Region-X. The aim is to sustain the delivery of relevant, timely and quality testing services and realize benefits for enterprises in terms of product safety or quality, increased income or access to new markets, employment generation, product development and compliance to regulations. The evaluation used interrupted time series quasi-experimental and descriptive designs, gathered primary and secondary data and utilized quantitative and qualitative techniques to obtain empirical evidence to determine project effectiveness. Simple and multiple linear regressions with interaction analyzed the data.

Keywords

program theory driven evaluation; counterfactual; interrupted time series; moderator; testing services evaluation; DOST

Introduction

Project designers normally base an intervention on a set of explicit and implicit assumptions about what actions are required in order to solve a problem. This set of assumptions is the program theory (PT) Chen (2005) described program theory as “a specification of what must be done to achieve the desirable goals, what other important impacts may also be anticipated, and how these goals and impacts would be generated” (p.16). Evaluators understand program theory as the underlying assumptions of a policy, of a project, or of an intervention in general. PT driven evaluation is also known as PT based evaluation (PTE) (Weiss, 1997).

PT driven evaluation starts with explication of the program theory. The formulation of causal relationships between variables representing the assumptions follows. Then, empirical testing confirmed or disconfirmed the causal hypotheses. This is testing the program theory in the literature of theory driven evaluation. When an evaluation is by the use of experimental or quasi-experimental designs, and includes a counterfactual estimation, it is an impact analysis according to World Bank evaluation standards (Baker, 2000; Kapoor, 2002; Leeuw & Vaessen, 2009).

This paper has five sections. First, it presents an overview of the project and the rationale of the study. Second, it illustrates the PT of the RURL project. Third, the article discusses the conceptual framework, evaluation paradigm, strategy, research methods and its limitations. Fourth, it presents the results. Finally, the paper identifies implications and subsequently proposes recommendations.

¹Department of Science and Technology-X, Carmen, 9000 Cagayan de Oro City
Email: melrats@yahoo.com

The Project and the Evaluation Rationale

RURL project's intended beneficiaries are the enterprises. The enterprises are of central interest to DOST because it is the basic organizational unit that utilizes scientific and technological knowledge to create products and in so doing effect economic and social benefits to the people. Enterprises contribute to the creation of wealth, employment, and income generation, both in rural and urban areas. As of 2009, there are 28,553 business establishments in Northern Mindanao with 91.68 percent (26,176) micro, 7.72 percent (2,204) small, 0.33 percent (94) medium and 0.25 percent (79) large. Of the 780,437 enterprises in the country as of 2009, 14 percent are in the manufacturing sector and 47 percent are in the food and beverage subsector (<http://www.dti.gov.ph>).

The provision of support to the enterprises is the fourth strategy identified in the National Science and Technology Master Plan for 2002-2020 (NSTP). Testing services through the operation of the RSTLs is one of the core programs under this strategy. DOST offered the services since stakeholder needs assessment identified that a strong support in terms of analytical testing would be crucial to the operation of the enterprises especially in the arena of product competition that is brought about by World Trade Organization (WTO) and similar agreements.

Common problems resulting from limitations of workforce, facilities and equipment beset all the regional offices that had established the testing services. Items of equipment were deteriorating and budget for replacement and repair was limited. There were only three (3) analysts of DOST-X. From surveys, focus group discussions and workshops conducted, DOST realized of the various unmet testing needs of the enterprises. Building and upgrading of Science and Technology (S&T) infrastructure is the sixth strategy identified in the NSTP and it supported the decision to implement the project in May 2007 with scope limited to chemical and microbiological testing. The project's duration is four (4) years and it is now on its fourth year. At the end of the program, government will have spent a total of P156,031,906 and about P9,254,000 for DOST-X. Items of appropriate equipment were acquired, facilities and systems were upgraded to international standards, three job order analysts were hired, trainings were provided and benchmarking activities were done in a rationalized approach.

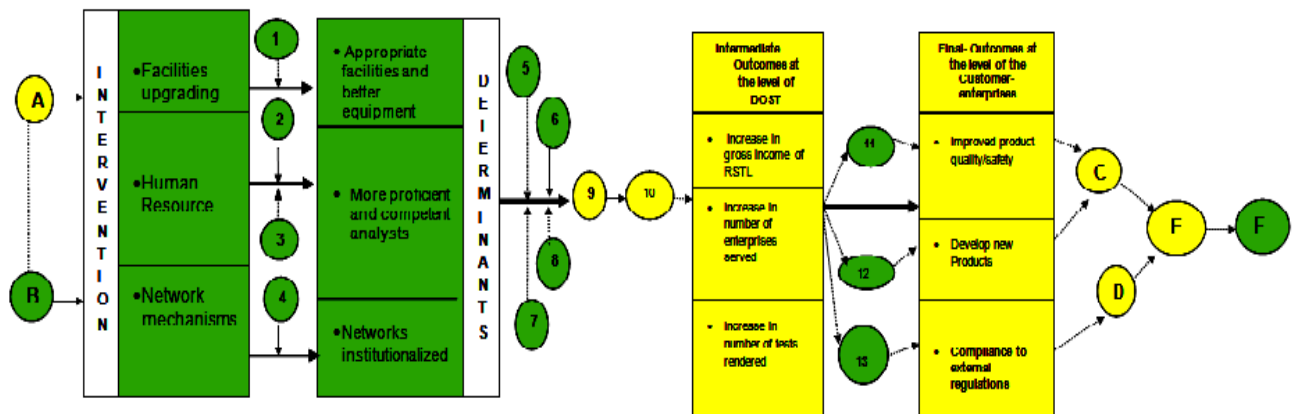
The project should have ended in May 2011 but was extended until December 31, 2011 and it is just cogent to find out what is the impact of the project. Was the testing services delivered well? Was the program theory of the project valid? Did the project accomplish the desired outcomes? Is the service worth sustaining, or should a new project with the same program theory be implemented? This study tried to find answers to these and other questions.

The RURL Program Theory

Figure 1 presents the chain of explicit (stated in project documents) and implicit (generally understood by project designers) assumptions of the project as revealed by project documents, from consensus and focus group discussion involving project planners, managers, coordinators and implementers (Leeuw, 2003). RURL provided facilities and equipment upgrading, workers augmentation, capability upgrading, and institutionalization of mechanisms for networking. The idea was that these components would produce efficient and reliable equipment, appropriate facilities, proficient or competent analysts and mechanisms for networks. Together with the implicit assumptions will enable DOST to sustain the delivery of timely, relevant and quality services. The implicit assumptions are the implementation and maintenance of ISO/IEC 17025 accreditation, top management and contextual support in the form of timely

release of funds, efficient procurement process, regular monitoring and internal evaluation by PCIEERD and accreditation or recognition by regulatory agencies, provision of incentives to implementers and action on customer feedback.

Consequently, the enterprises can now avail of more testing services. Hence, the project would produce the desired outcomes of increases in the number of tests conducted, of enterprises served and gross income of the Regional Standards and Testing Laboratories (RSTL). The more tests an enterprise availed will provide more information to help it improve the quality and safety of its products, develop new products and comply with regulations. Finally, these will lead to employment and socioeconomic benefits to the people and national development.



Legend:

A-Enterprises need varied testing services for purposes of product safety and quality assurance, product development and compliance to regulations. Many enterprises do not have their own laboratory and those who have still need third party validation.

B-Before 2007, DOST cannot provide most of the needed tests because of limitations in workers, facilities and capabilities

1-Replace irreparable equipment, provide new equipment and appropriate facilities so that more tests will be available

2- Augment workforce so that more tests can be done

3- Provide relevant trainings to improve proficiency and competence of analysts

4-Build networks, facilitate ad support referral system and complementation

5-Implement and maintain ISO/IEC 17025 accreditation

6-Give strong top management support in terms of timely provision of funds and efficient procurement process

7-Provide incentives to implementers

8-Action on customer feedback

9-Sustain the delivery of timely, relevant and quality services

10-Enterprises avail more services

11-More tests availed by an enterprise will provide more information to help it improve the quality/safety of its products

12-More tests availed by an enterprise will provide more information to help it develop new products

13-More tests availed by an enterprise will help it comply with regulations

C-Improved or maintained product quality/safety and new products will expand markets and increase income and thereby employ workers

D-Compliance to regulations would enable an enterprise to operate, earn income and hence continue to employ workers

E-Employment is generated

F-Employment will bring socio-economic benefits to the people and national development.

Note: The research gathered empirical data on the assumptions in yellow.

Figure 1. The Chain of Assumptions of the RURL Project

The Conceptual Framework

The **Program Theory Driven Model of Evaluation** of Chen provided the roadmap for this evaluation study (Figure 2) while the basis of the program was the needs of the stakeholders or stakeholder theory (Chen, 2005). Two theories or models imparted the logic; the **Model of Sufficient Component Cause** proposed by Rothman in 1998, (Rothman, Greenland & Lash, 2008) and **Counterfactual Model** for causal analysis of Rubin often referred to as the **Potential Outcomes Framework** (as cited in Morgan & Winship, 2009). VanderWeele and Robins (2006) wrote that Rothman's theory of sufficient component cause is notably associated with Mackie's *inus* condition.

INUS means “an insufficient but nonredundant part of an unnecessary but sufficient condition” (Shadish, Cook & Campbell, 2002 p.4). A cause that is an *inus* condition is an insufficient cause by itself, and its effectiveness required it to be one of the components of a larger set of conditions. Most causes are actually *inus* conditions. Many factors are usually required for an effect to occur but these and their relationships are not known completely (Shadish, Cook & Campbell, 2002). The counterfactual or potential outcomes framework is giving an account of the effects of particular causes or interventions. The theory of sufficient component cause is giving an account of the causes of particular effects.

Chen’s PT framework is composed of the action model and change model. The framework says that the implementation of the program starts with acquisition of resources from the environment and this activates the action model. The action model has six elements but some programs have less but just as effective. The activation of the action model leads to the implementation of the program and initiation of the change model. The change model elements of intervention, determinants and outcome have causal relationships. Outcomes can be intermediate or final and can be at the level of the implementing institution and at the beneficiaries (Chen, 2005). Setting an outcome at the level of the implementing institution is logical and facilitates internal evaluation. Feedback channels are shown as arrow headed dashed lines in figure 2.

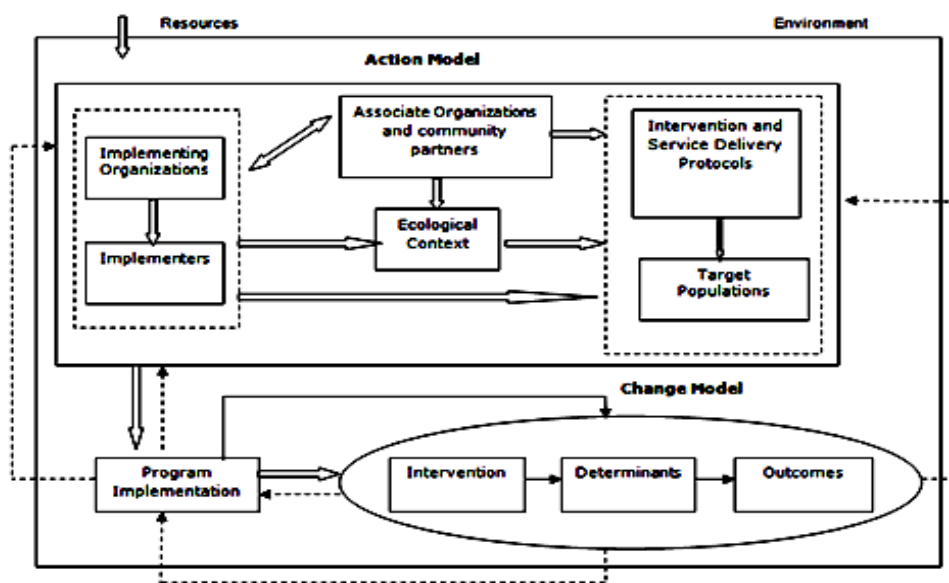


Figure 2. Program Theory Framework of Chen (2005).

The Evaluation Paradigm

Figure 3 shows the change model of the project RURL based on the chain of assumptions of RURL shown in figure 1. The determinants are actually what were absent that is why DOST RSTLs could not provide most of the testing services needed by the enterprises. The project activities shown in a Gantt chart are towards the creation of the determinants. Increases in income of RSTL, in the number of enterprises served and in the number of tests rendered are intermediate outcomes because these can only result from the synergy of the determinants and quantifiable only over a period of time and after implementation of technical as well as administrative processes.

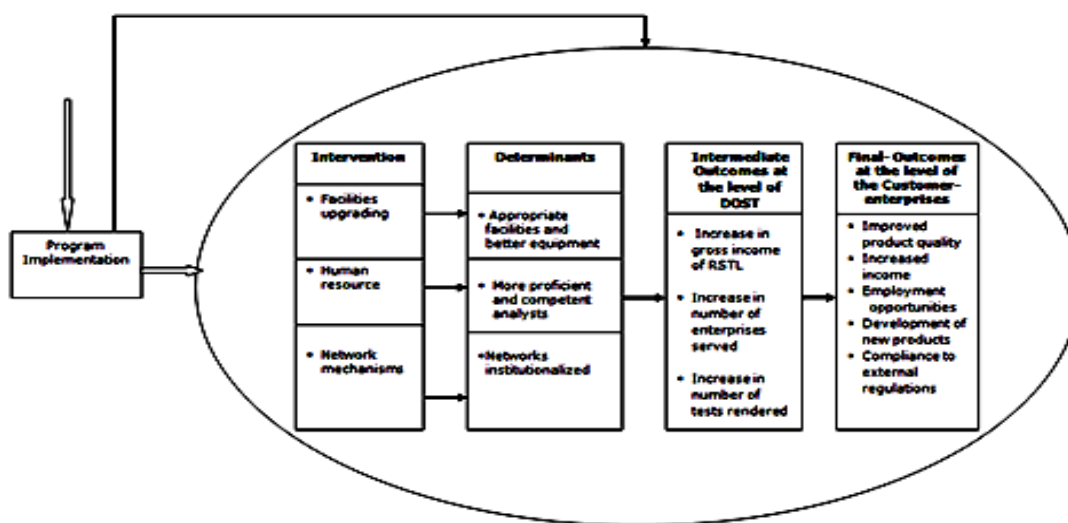


Figure 3. Change Model of the RURL project

Figure 4 is the schema that shows the variables of the study (except for program implementation). It shows two causal relationships. The intervention and determinants are put together as one box because these were already the subject of internal monitoring and evaluation, hence excluded in this study. Thus, the research paradigm posited that the intervention components provided sufficient cause for the determinants to emerge just as the determinants provided sufficient cause for the potential outcome to appear as the logic of the theory of sufficient-component cause argued. The first causal link to be studied is thus from determinants to intermediate outcome. The set of three determinants are the component causes and together with some other causes not the subject of this study are sufficient to bring about the intermediate-outcome according to the theory of sufficient component cause. The evaluation of the intermediate-outcome is a scenario where the particular events or causes are given and the research estimated the effects or potential outcomes. VanderWeele and Robins (2006) characterized this scenario under the counterfactual or potential outcome framework. The causes are the determinants and the effects are the intermediate-outcomes of the project. The intermediate outcome is in terms of any of the three variables. These three are indicators of the same event. The use of the three is to support internal validity of the chosen research design. Increase in income of RSTL means the increase in the amount of fees collected from all the customer-enterprises from the conduct of tests.

For the second causal link from intermediate to final outcome, only one indicator was employed as the independent variable and this is the volume of patronage in terms of the annual average number of tests requested per customer-enterprise from 2004-2010. This choice of the independent variable was necessary because the outcome variables are also in terms of the effects of the project as perceived by each customer-enterprise so that the relationship would be on the same DOST unit of analysis. The inclusion of intermediate to final-outcome as another causal process linked the DOST testing services as one component cause and the desired outcome for the enterprises as the effect.

The final-outcomes at the level of the customer-enterprises consisted of five indicators given in the change model. This scenario is the question: given a particular

effect, what are the various events that might have been its cause? The research paradigm assumed that component-causes or events that are in place are sufficient so that the survey measured the extent of the effects even at perceptual basis only. Further, the causal relationship situated the effects as the final-outcomes and the cause was the testing services of DOST.

Improved product quality is the respondents' perception about the quality of their enterprise's products because of the testing services. Increase in income pertains to the gross product sales of the enterprise which resulted from an increase in production because of the increase in demand resulting from improved quality of products. Employment opportunity is the need or capacity of the enterprise to offer more person-hours of employment. Development of new products pertains to the ability of the enterprise to create variants or entirely new products. Compliance to regulations is a situation where the law allowed the enterprise to operate because the requirements of such were met by them.

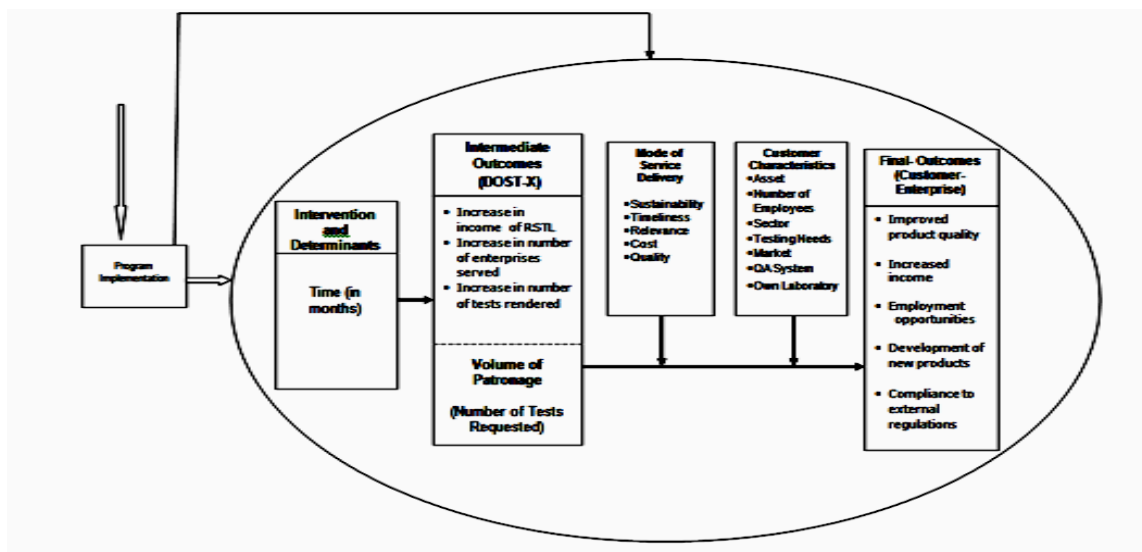


Figure 4. The Variables of the Study

The Evaluation Strategy, Methods and Limitations

The first causal link was tested by interrupted time series design. The independent variable is time (months) and the dependent variables are the intermediate-outcomes. The second causal link was tested using the dose-response approach (Sechrest and Yeaton, 1982). The dose is the independent variable in annual average number of tests availed and the dependent variables are the final-outcomes.

Service delivery and customer characteristics were moderators or control variables in the analysis of the second causal link and each analyzed one at a time as a two-way interaction. The moderator is a factor used to search for co-variation and test for non spuriousness (O'Sullivan, Rassel & Berner, 2003). The joint action of component causes may be one of interactions (VanderWeele & Robins, 2006). Multiple linear regressions with the introduction of an interaction capture the statistics of this theory of sufficient cause (Pearl, 2009). The introduction of interaction variable and the test of significance of the interaction coefficient in the analysis is a technique of testing an assumed causal relationship. Service delivery in terms of the objectives of the project is described as sustainable, timely, of reasonable cost and high quality. Quality was measured in this study in terms of the five dimensions of SERVQUAL developed by

Zeithaml et al. in 1990 (as cited in Nyeck et al., 2002); reliability, assurance, tangibles, empathy and responsiveness. Enterprise categories are asset, number of employees, sector, testing needs, market, presence or absence of a quality assurance system and ownership of a testing laboratory.

The 11 specific questions tacked in the study can be grouped into four according to the purpose. First is the explication of the program theory and the resulting framework is given in Annex A. The second set of questions is on the context of the program and designed to identify moderators and explain variance in the result. Questions are on the profile of enterprises and their perceptions on the delivery of the testing services. The third set is the interrupted time series design that will answer the causal link from determinants to intermediate-outcome with the logic of counterfactual or potential outcome framework. Time series regression, forecasting with Holt's trend corrected exponential smoothing (Bowerman, O'Connell & Koehler, 2005) and multiple linear regression with interaction of the form, $Y_t = \alpha + \beta X_t + \beta_1 T_t + \beta_2 X_t T_t$ (Mohr, 1995) were used. Counterfactual and point impact on May 2009 after two years of implementation were estimated by projection of the value of the intermediate outcomes using the regression equation modeled for the before intervention. Counterfactual is the value of the variable in the absence of an intervention (Shadish et al., 2002). Impact is the difference between the actual value with the intervention less the counterfactual (Mohr, 1995). The fourth set is an exploration of the causal link from intermediate to final outcome. Descriptive design was used with the moderating mechanism approach. This anchored on the logic of theory of sufficient component cause. Simple linear regression ($y = \alpha + ax$) and the multiple linear regression with interaction in the form, $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$ (Weinberg and Abramowitz, 2008) were used.

Project designers, monitors and implementers responded to questionnaire I that was circulated by email to all the elements in the sampling frame. Sixteen out of 18 people (89%) responded. Philippine Post circulated questionnaire II to all enterprises that have availed of the testing services of DOST-X by priority mail with self-addressed and stamped return envelope. Questionnaires were sent on November 15, 2010 and the last response counted was in by January 19, 2011. Fifty two out of 237 (22.5%) responded even with two rounds of follow-up by telephone.

Counterfactual and impact estimation was limited in terms of testing fees collected by the DOST-X, the number of enterprises served, and the number of testing services rendered after two years of implementation. The Independent Evaluation Group (IEG) of World Bank and the International Initiative for Impact Evaluation (3ie) suggest that rigorous evaluation designs such as the experimental and quasi-experimental designs must be employed in doing impact evaluations especially if these are development programs (Ding, 2006). In this study, the absence of a control group and random selection excluded the use of experimental design. Here is a case of self-selection wherein the population who decided to avail of the services, received it. The use of more than one outcome indicator compensated for the lack of a control group. The results of the study revealed the lack of statistical power in this self selection survey type research.

The moderators were limited to service delivery and some characteristics of the enterprise. The conceptual framework of the study assumed linear and unidirectional relationships. The regression models were deterministic. Furthermore, the interrupted time series models neglected the presence of autocorrelations because these were not used to forecast but only for description. Finally, in terms of depth in the study of causal relationship of number of testing services and extent of final-outcome, this part of the

research is just exploratory in nature because the measures of effects were only perceptions. Moreover, the study did not produce a complete causal description and explanation because it was limited to a few independent variables as moderators. Finally, although the scope of this government project was nationwide, this outcome evaluation was only limited to Region-X.

Results

The PT of the RURL project in Chen's framework is shown in Annex A. Customer-enterprises rated the overall service delivery of DOST-X RSTL as good with a weighted mean of 3.09 on the scale of 1-4. The dimensions of quality, relevance, timeliness, sustainability and cost have means of 3.38, 3.39, 3.09, 3.00 and 2.96 respectively. There were no significant differences between the respondent-enterprises' rating on the delivery of service when grouped into categories of asset, number of employees, sector, market, presence of quality assurance system and having their own laboratory.

The distribution of the customer enterprises of DOST-X RSTL in terms of number of employees category were, 29 percent micro, 31 percent small, 15 percent medium and 25 percent large enterprises. In terms of asset category, 29 percent was micro, 31 percent was small, 15 percent was medium and 25 percent was large enterprises. In terms of sector or industry classification, 37 percent was in food and beverage processing, 32 percent was in water services (purification and distribution), two percent was in health services and restaurants, eight percent was in agriculture and forestry and 23 percent was in non-food manufacturing and construction. About 2/3 of the enterprises (69%) supply their products and services only to the local market and about 1/3 (31%) are exporters. Ninety-five percent of the enterprises need chemical and microbiological testing services and about five percent need other services as well. Twenty-one percent of enterprises need microbiological testing only while 12 percent need chemical testing services only. Majority of the enterprises have a quality assurance system (85%) while only almost half (46%) own a testing laboratory.

According to the perception of the respondents, on the scale of 1-10, the extent of outcome or extent of effect or indirect help of the testing services to their enterprises was big contribution, a mean of 7.58 in terms of compliance to regulations. In terms of quality of products, income to the enterprise and development of new products, the means were respectively 6.93, 6.00 and 4.98 described as medium contribution. In terms of employment, the rating was small contribution with a mean of 4.55. The study also revealed that customer-enterprises regard six factors that caused the given desired effects. In order of importance, these are; quality assurance policy, good management, high demand of products, testing services of DOST, requirement of buyers and bright economy. Seventy three percent of respondents said that the testing service is a factor along with good management.

There were significant differences between the trends in income to RSTL, number of enterprises served and number of tests rendered before and after the implementation of the project. The trends were decreasing before the project and these shifted to increasing after project implementation. The plots showed a definite change in pattern after six months of implementation by visual inspection starting 2008. There was no seasonality and fluctuations were high with very low points before, during and after PAO audits. The estimated counterfactual on May 2009 after two years projected from May 2007 at the start of project was P21,044 for income to RSTL, nine (9) for enterprises served and 31 for tests rendered. On May 2009, the approximated point

impact was P94,514 in increases in income of RSTL, 21 more enterprises served and 206 more tests rendered.

There were no significant relationships between volume of patronage and extent of final-outcome in all five indicators. However, there were significant relationships between volume of patronage and the final-outcome indicators of income, development of new products and employment with service delivery as moderator. This means that for the same volume of patronage, the extent of outcome in the three indicators mentioned are greater when the service delivery is better. There was no significant relationship between volume of patronage and extent of outcome in terms of quality of products or compliance to external regulations with service delivery as moderator. Volume of patronage does not predict the extent of outcome in quality of products and compliance to regulations even at different levels of service delivery. Table 1 shows the summary of results.

Table 1. Summary of Results of Linear Regressions with Volume of Patronage as Independent Variable 1, Service Delivery (Moderator) as Independent Variable 2 and their Interaction as Independent Variable 3 ($y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$)

Dependent Variable of Extent of Effect on	Multiple Correlation Coefficient (<i>R</i>)	Significance of Model, <i>p</i> -value	R ²	Significance of Coefficient of Interaction Variable β_3 , <i>p</i> -value
Quality of products	0.36	0.12 ^{ns}	0.13	0.08 ^{ns}
Income	0.46	0.02*	0.21	0.009*
Employment opportunities	0.48	0.02*	0.23	0.015*
Development of new products	0.490	0.02*	0.20	0.03*
Compliance to external regulations	0.18	0.60 ^{ns}	0.05	0.51 ^{ns}

* Significant

ns-not significant

There was a significant relationship between volume of patronage and extent of outcome in terms of quality of products with enterprise category of MSME or not as moderator. Volume of patronage significantly relates to the extent of outcome in quality of products for MSMEs but not for large enterprises. There was no significant relationship between volume of patronage and extent of effect in compliance to external regulations with any of the tested moderators of enterprise characteristics.

Conclusions

The interrupted time-series part provided empirical evidence that the project caused the desired intermediate outcomes. NSCB reported that the GRDP rate of Northern Mindanao in 2007 was 7.9 percent and in 2008, it decreased to 5.3 percent (<http://www.nscb.gov.ph>). There was no other event in history that could possibly be a cause of the increases in intermediate-outcomes other than the project.

The descriptive design part revealed that the DOST's testing services had an effect on the quality of products, income, employment, development of new products and compliance to regulations of its customers. The results showed that the project accomplished the desired outcomes on the part of the customer-enterprises. Customer-enterprises rated the delivery of service and quality as good. DOST-X RSTL delivered

the testing services fairly across categories and sub-categories of customers. The study therefore concludes that the project RURL is effective.

The research hypothesized that there will be significant correlation between volume of testing request and extent of final-outcomes. The program theory assumed it. However, the results showed no significant relationship in all indicators. The failure of the study to reject the null hypothesis could be due to three major factors; statistical power, sensitivity of the measure for final-outcome indicators and that the null hypothesis is in fact true. Nonetheless, the study concludes that volume of patronage is not a significant predictor of extent of final-outcome indicators.

For the same volume of patronage, the extent of outcome in quality of products, income and development of new products would be higher, the better the service delivery. The effect of the testing services was bigger for MSMEs than to large enterprises. This study did not find a significant moderator of the relationship between volume of patronage and extent of outcome in compliance to regulations given the factors included. Quantitative data and statistical inferences completely proved the first causal link. Descriptive statistics, qualitative information, quantitative data, statistical inferences and theory only supported the second causal link. Nonetheless, the study concludes that the program theory of the project is valid. However, further studies must be conducted to completely describe and explain causal link 2.

Results on the profiles of enterprises and service delivery of DOST-X RSTL provided data to make confirmation about some elements in the action model of the program theory. Results of good service delivered fairly to correct target population provided empirical evidence on the two elements of service delivery protocol and target population. From the results of the profiling, the study concludes that DOST-X RSTL had served only a negligible fraction of the population of micro enterprises and a small fraction of the population of small enterprises but a bigger fraction of the medium and large enterprises.

Implications and Recommendations

The study concluded of significant differences in trends in intermediate-outcome indicators before and after the project. However, as to whether these significant increases in trends are meaningful would call for a cost and return analysis vis-à-vis the organizational values and goals of DOST.

Computing from DTI data, there is about 2500 potential customers of DOST-X RSTL since the testing services were designed for the food, beverage and water sectors. With only 262 customers in its database, it means that DOST-X was able to serve only about 10 percent of the potential customers. This implies that it should try to reach out to more customers especially the micro enterprises. This also means that there would still be a need for this kind of intervention in the near future.

Therefore, the study proposes short term and long-term recommendations as well as suggestions for further research. Short-term recommendations are for the DOST-X office, long-term recommendations are for the DOST and government.

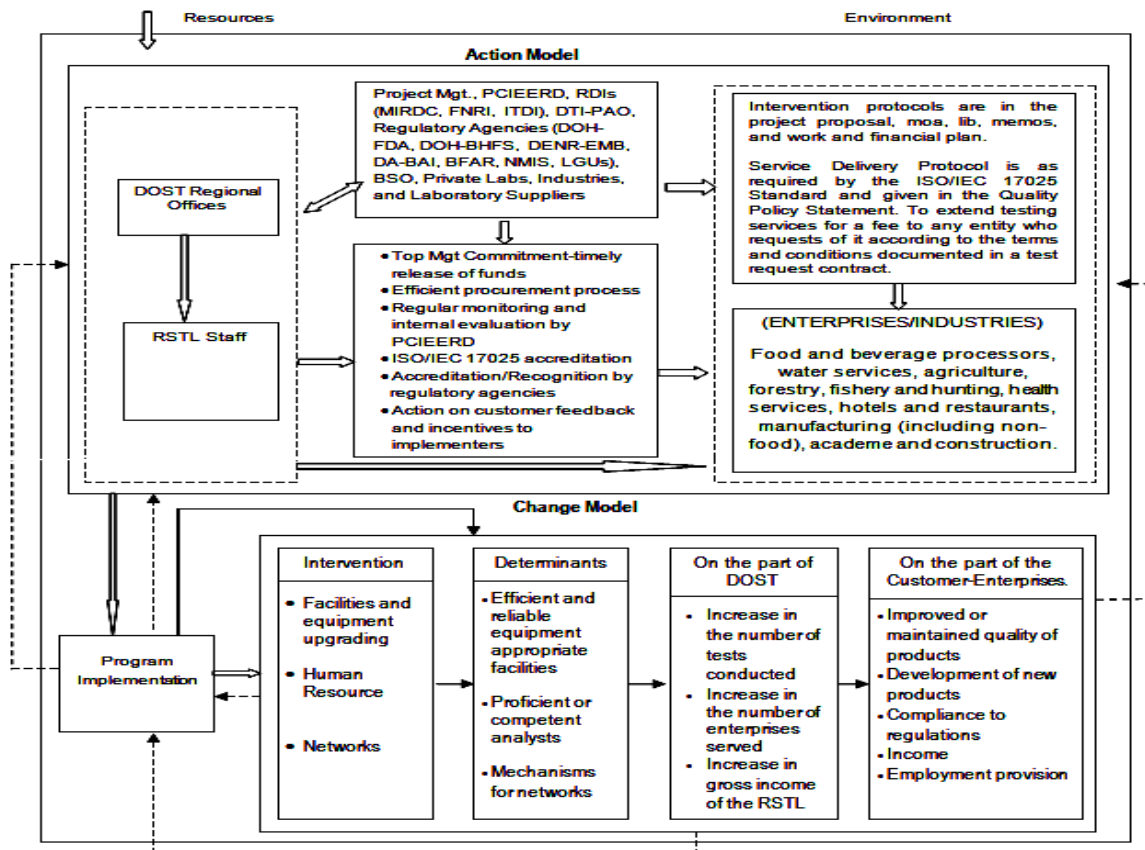
In the short term is improvement of the delivery of service in the areas that the study results pointed out. The study recommends to DOST-X the following: (1) simplify the forms and computerize sample receiving in order to improve the speed of processing of service request. Reduce transaction time to say 15 minutes. (2) Conduct a lecture-seminar on SERVQUAL quality dimensions to RSTL staff. This will improve their understanding of the expectations of a customer about quality service. (3) RSTL includes in its database a profile of the enterprises. Office validates and updates

information such as complete mailing address and telephone numbers as well. This will facilitate feedbacks. (4) RSTLs database other laboratories (private and government) and their services so that it may refer un-served customers to them. On empathy, the problem whether the office gives referrals or links customers to other laboratories got the lowest rating. (5) More regular positions to be given to chemists and microbiologists. (6) Provide more working hours in microbiological testing where waiting time is longer and demand is higher. (7) Reach out to the micro enterprises. Promote the testing services to them.

In the long-term, the study recommends to government the following: (1) Policy on the inclusion of a PT in all types of interventions. The PT framework must be a part of the proposal because it facilitates program planning, implementation, monitoring, and evaluation. (2) Policy that encourages wide discussion by stakeholders including evaluators, social scientists and other theorists in the process of conceptualization and implementation in order to come up with valid program theory and increased effectiveness. (3) Assess service delivery in every project evaluation. The results of the study showed the importance of good service delivery for the effective implementation of programs.

To researchers and evaluators, the study recommends the following: (1) Testing of the hypothesis that there is no significant relationship between volume of patronage and extent of final-outcome used in this study with a sample size of more than 250 to increase statistical power. More test factors may be included. (2) Design evaluation studies that search for sensitive yet simple final-outcome indicator measures. (3) Further cost and return analysis with the quantitative results from this research. (4) Conduct program theory driven evaluations of more government programs.

Annex A



Program Theory of "Rationalized Upgrading of DOST RSTLs" in Chen's Framework

Acknowledgments

The author discloses receipt of financial support from the DOST Human Resource Development Program in the conduct of this study. She is grateful to DOST and Xavier University-Ateneo de Cagayan Graduate School.

References

- Baker, J. L. (2000). Evaluating the Impact of Development Projects on Poverty: A Handbook for Practitioners. *Directions in Development*, Washington D.C.: World Bank. <http://siteresources.worldbank.org/INTISPMA/Resources/Impact-Evaluation-Handbook-English-/impact1.pdf>.
- Bowerman, B.L., O'Connell, R.T., & Koehler, A.B. (2005). *Forecasting, Time Series, and Regression*. 4th ed. 2005. Australia: Thomson Brooks/Cole.
- Chen, H. T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. Thousand Oaks, CA: Sage.
- Ding, C. S. (2006). Using Regression Mixture Analysis in Educational Research. *Practical Assessment Research and Evaluation*, 11(11), Retrieved January 18, 2010, from Proquest: <http://library.xu.edu.ph/index.htm>. 2006.
- DOST (2007). Rationalized upgrading of DOST Regional Standards and Testing Laboratories. *Project Proposal*.
- Kapoor, A.G. (2002). Review of Impact Evaluation Methodologies Used By the Operations Evaluation Department over Past 25 Years. Washington D.C.: World Bank. <http://www.worldbank.org>.
- Leeuw, F. & Vaessen, J. (2009). Impact Evaluations and Development: NONIE Guidance on Impact Evaluation. Washington D.C.: World Bank, 2009. <http://www.worldbank.org/ieg/nie>.
- Leeuw, F. L. (2003). Reconstructing program theories: Methods available and problems to be solved. *American Journal of Evaluation*. 24 (1) 5-20. <http://www.eval.org>.
- Meier, K. J., Brudney, J.L., & Bohte, J. (2009). *Applied statistics for public and nonprofit administration*. Belmont, CA: Thomson Wadsworth.
- Mohr, L. B. (1995). *Impact Analysis for Program Evaluation*. 2nd ed. Thousand Oaks, CA: Sage.
- Morgan, S.L. & Winship, C. (2009). *Counterfactuals and causal inference: Methods and principles for social research*. Excerpt 978-0-521-85615-7. Cambridge University Press. <http://www.cambridge.org>.
- Nyeck, S., Morales, M., Ladhari, R., & Pons, F. (2002). 10 years of service quality measurement: reviewing the use of the SERVQUAL instrument. *Cuadernos de Diffusion*, 7(13), 101-107. <http://library.xu.edu.ph/index.htm>.
- O'Sullivan, E., Rassel, G.R. & Berner, M. (2003). *Research methods for public administrators*. 4th ed. New York: Addison Wesley Longman, Inc.
- Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, 3, 96-146. <http://www.i-journals.org/ss>. DOI: 10.1214/09-SS057.
- Rothman, K.J., Greenland, S., & Lash, T.L. (2008). *Modern Epidemiology* 3rd edition (Preview). Lippencott Williams and Wilkins. <http://books.google.com.ph/books>.
- Sechrest, L.B., & Yeaton, W.H. (1982). Magnitudes of experimental effects in social science research. *Evaluation Review*. 6, 579-600. <http://www.eval.org>.
- Shadish, W. R., Cook, T.D., & Campbell D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin Company.
- VanderWeele, T.J. & Robins, R. M. (2006). *A theory of sufficient cause interactions*. Berkeley: Electronic Press. 2006 retrieved November 7, 2010. <http://www.biostats.bepress.com/cobra/ps/art113>.
- Weinberg, S.L. & Abramowitz, S.K. (2008). *Statistics using SPSS: An integrative approach*. 2nd ed. New York: Cambridge University Press.
- Weiss, C. H., (1997). How can theory-based evaluation make greater headway? *Evaluation Review*. 21(4), 501-524. <http://erx.sagepub.com>. DOI: 10.1177/0193841X970210045.