

CHAPTER 1: USING LOGIC MODELS

Learning Objectives

This chapter will help students gain an understanding of the following:

- Rationale for logic models
- Elements of a basic logic model
- Process for developing a logic model
- How to avoid pitfalls of logic modeling.

Key Terms

- Resources
- Activities
- Customers
- Outputs
- Outcomes
- Contextual Factors
- Program Theory
- Antecedent Variables

Review Questions

1. In what ways does the “if/then” approach to logic models clarify the program processes?
2. In what ways can mediating factors positively and negatively affect an evaluation? Why should evaluators be aware of them?
3. Describe some benefits resulting from gaining stakeholder agreement on the program logic before designing the research questions or the means of data collection.

Discussion Questions

1. Describe the elements of a basic logic model.
2. Identify the advantages and disadvantages of using a logic model.
3. What are some methods for overcoming the disadvantages?
4. When using a logic model as part of the evaluation process, who should participate in building the model?
5. Compare some different situations in which a logic model might be used, and discuss how participation in the model building might be affected.

Exercises

1. Head Start, a program funded by the U.S. Department of Education, focuses on enabling low-income, disadvantaged children to attend preschool for the purpose of better preparing them for future educational endeavors. There has been a movement within Congress to eliminate this program, because studies have shown that the program is not achieving its intended outcomes. You work in the office of a member of Congress whose constituents are low-income and have received direct benefits through this program. You have been asked to serve on a steering committee for an evaluation being conducted of the local program. The intention of this evaluation is to give your member of Congress a tool to help buffer the criticisms of Head Start.

In small groups, demonstrate your understanding of the logic model by putting together a logic model including all relevant aspects (listed in Key Terms,) centered on the Head Start Program.

2. You have been asked by Homes Against Homelessness, a nonprofit organization that operates programs assisting people without homes, to conduct an evaluation for presentation to philanthropists who have donated to the nonprofit. Being an expert in the program evaluation field, you suggest to the nonprofit that a logic model could provide a demonstration of what the nonprofit does and how it impacts the community. Explain to the nonprofit the steps taken to create a logic model. Include an explanation of how each step of logic model provides insight and purpose for each part of the organization. Create a possible logic model for the organization based on activities such as providing shelter, food, clothing, job training, and child care to those without homes.

Contributors: Rachel Wyatt and Patrick Wells

CHAPTER 2: EVALUABILITY ASSESSMENT

Learning Objectives

This chapter will help students gain an understanding of the following:

- Useful outcomes of evaluability assessments
- Steps of the evaluability assessment process
- Elements of a successful assessment
- Overcoming obstacles to a successful assessment

Key Terms

- Evaluability
- Program intent
- Program goal plausibility
- Program reality
- Evaluation priorities
- Evaluation criteria
- Alternative evaluation design

Review Questions

1. When is it appropriate to conduct an evaluability assessment? When is it not appropriate? Compare the two types of circumstances.
2. What are the six key steps of an evaluability assessment? Who should be involved in the assessment?
3. What factors should be considered when identifying evaluation design criteria? Why?
4. Identify problems that might inhibit useful program evaluation. What are some approaches for overcoming these problems?

Discussion Questions

1. If an assessment seems to indicate that a program's goals are implausible, what steps might an evaluator take in concert with the program managers to address this?
2. Why is it important that the evaluator and the policy makers or program managers agree on the design of the evaluation ahead of time? What are the implications of not discussing and agreeing on the design in advance?
3. How might cost be a factor in an evaluability assessment? Discuss methods for reducing the costs of an assessment and subsequent evaluation.

Exercises

1. The State of Maryland funds a program intended to help children in foster care who are reaching the age of eighteen transition smoothly into independent living after they "age out" of the foster care system. The program is operated by local contractors, such as community colleges, who provide an educational program for older teenagers on life skills they will need to function as adults. These contracted service providers coordinate with the local state welfare offices to work with teens who are referred to the program, which is voluntary. The state would like to know whether the program is effectively helping teens make the transition out of foster care. Describe why an evaluability assessment might be appropriate and what steps you would take to accomplish such as assessment. Be sure to include all interested parties in the assessment, as appropriate.
2. A large city is revamping its public housing program. It has begun demolishing large, older, outdated housing projects and relocating the residents to dispersed locations throughout the city where low-income housing is interspersed with more expensive housing units. The city is provided a wide array of services to the families being relocated and would like to conduct an evaluation to ascertain the effectiveness of the services and their delivery. Plan and outline the steps you would take to conduct an evaluability assessment for this activity. Describe some of the problems one might expect to encounter and approaches to overcoming these problems.

Contributor: Nancy A. Potok

CHAPTER 3: IMPLEMENTATION EVALUATION

Learning Objectives

This Chapter will help students gain understanding of the following:

- Relationship between program implementation and implementation evaluations
- Black box paradigm versus transparent box paradigm
- Strategies and tools for conducting various types of evaluations
- Four stages of chronological evaluation

Key Terms

- Black box paradigm
- Transparent box paradigm
- Formative evaluation
- Summative evaluation
- Process evaluation
- Descriptive evaluation
- Need and feasibility assessment stage
- Program planning stage
- Program delivery stage
- Implementation research review
- Program improvement stage

Review Questions

1. Compare and contrast between formative, summative, process, and descriptive evaluations.
2. What are the four stages of chronological evaluation? Specify their main characteristics.

3. Describe the benefits and challenges of an implementation research review. How would such a review complement key informant interviews?
4. What are the benefits and challenges of using coverage analysis, component analysis, program records, and case studies to evaluate program delivery?
5. Describe the relationship between program logic models, program templates, and outcome hierarchies. When would it be appropriate to use them?

Discussion Questions

1. How would you ensure staff participation to prepare program logic models, program templates, and outcome hierarchies in a human rights advocacy organization?
2. How would you develop coverage analysis for a drug prevention program in neighborhoods at risk?
3. Describe the risks of using client feedback for a teenager early pregnancy prevention program.

Exercises

1. An international organization is implementing an environment productivity project in a rural coastal village of three hundred people in a Central American country. People in the village are poor, living at subsistence level in a rural area with little support from the government. The objective of the program is to help villagers improve their standard of living by teaching them land cultivation techniques while protecting the environment and producing income. Environmentally friendly practices for agriculture as well as business principles need to be learned by the villagers. The program components include teaching villagers 1) how to cultivate their own land while protecting the environment and 2) productive activities such as selling herbal medicinal plants to nearby villages.

In a group discussion, decide how you would carry out a chronological implementation evaluation. Describe what type of evaluation would be best utilized for each stage, the potential risks and challenges regarding client participation, data-collection methods, management approach, and staff support. What are the outcomes expected from the program, and how will you monitor and measure the activities that lead to these outcomes?

2. An international NGO is implementing a gender-focused microfinance project in a rural village in Bangladesh. Most women in the village are household heads and need money to support their children and other family members. The women live from informal production activities such as selling woven handcrafts in nearby markets. Some women prepare meals to be sold in the marketplace. The objective

of the program is to help the village women improve their businesses. The components of the program are: 1) providing training on basic finance and accounting principles and 2) providing training on business management skills such as bookkeeping, improving sales, marketing, and managing capital.

In a group discussion, decide how you would carry out a chronological implementation evaluation. Describe what type of evaluation would be best utilized for each stage, the potential risks and challenges regarding client participation, data-collection methods, management approach, and staff support. What are the outcomes expected from the program, and how will you monitor and measure the activities that lead to these outcomes?

Compare your approaches to the two evaluations above, noting similarities and differences.

Contributor: Natalie Pazmino

Chapter 4: Performance Monitoring

Learning Objectives

This chapter will help students gain an understanding of the following:

- Identifying measures of program, agency, or system performance and providing information to managers and policymakers
- Developing formal goals, objectives, standards, and targets, program logic models, and the balance scorecard
- Appropriate data sources, operational measures, and quality assurance procedures in the performance design
- Comparing monitoring systems over time, actual performance against targets, comparisons among operating units, other kinds of breakouts, or comparisons against external benchmarks
- Importance of considering stakeholder involvement along with methodology

Key Terms

- Balanced scorecard
- Efficiency
- External benchmarking
- Productivity
- Program logic model
- Performance measures
- Outcomes
- Outputs

Review Questions

1. What are the different ways performance monitoring systems can track results?
2. What criteria should concern performance monitoring system designers in regard to the quality of their measures?

3. Name and describe the types of performance measures.
4. What are some comparative methods in which performance monitoring can be examined and used by managers?

Discussion Questions

1. In what ways do performance monitoring systems differ from evaluation studies, and what are some similarities?
2. Why would there be methodological concerns with determining what types of measures to use in a performance monitoring system?
3. Why are data from stakeholders important in performance monitoring systems, and in what ways can an evaluator ensure they are included?
4. Why is performance monitoring a useful tool for management and policy decision makers?

Exercises

1. Camp Great Outdoors operates eleven summer camps in Arkansas, Colorado, and Missouri. The camps have three sessions over the summer that each last twenty-six days. During that time campers participate in activities that include: football, basketball, baseball, soccer, tennis, rock climbing, rappelling, cliff jumping, water skiing, hiking, and spelunking (caving). With such dangerous activities involving children ranging from seven to eighteen years of age, safety is a primary concern of Camp Great Outdoors. Its motto is "Safety first, safety last, and safety in-between."

In order to lower liability insurance and improve the less-than-desirable safety record of the camp due to accidents that occurred last summer, Camp Great Outdoors is looking to develop a performance monitoring system to track the reported injuries and illnesses that result from its activities. The directors of the camp desire to know a breakdown of the injuries by session, severity, and activity. You have been hired to create a performance monitoring system for Great Outdoors before the first summer session starts in three months. Describe how you might develop a performance monitoring system that can pinpoint dangerous activities, thus lowering liability insurance payments and increasing overall safety for the campers.

2. Over the last decade, Birmingham, Alabama has experienced rapid population and economic growth. This growth has resulted in an increase in traffic congestion and the need to alter the city's long-range transportation plan. With the rapid changes taking place, the City Council feels the Traffic and Engineering Department should take measures to ensure everyone within the department is communicating. In order to achieve this communication, the Council recommends that the Traffic and Engineering Department create a balanced scorecard that will help align goals and measures of department heads with those at the lower levels of the department. You have been hired to help create a balanced scorecard for the Birmingham Traffic and Engineering Department. In doing so, identify what areas should be included in their performance measurement process.

Contributor: John Adam Almond

Chapter 5: Quasi-Experimental Design

Learning Objectives

This chapter will help students gain an understanding of the following:

- Differences between true experiments and quasi experiments
- Appropriate applications for using quasi experimentation
- Limitations of quasi experiments, including threats to the validity of quasi-experimental results
- Interrupted time series, comparison groups, nonequivalent group designs, and regression discontinuity designs
- Practical skills in using quasi-experimental design for program evaluation

Key Terms

- Quasi experimentation
- Pre- and post tests
- Internal validity
- Comparison group
- Nonequivalent dependent variable
- Cohort groups
- Quantitative assignment variable

Review Questions

1. Are quasi-experimental designs always less valid, or less able to be generalized, than random experimental designs? Why? Discuss the internal validity threats inherent to many quasi-experimental designs in your response.
2. Compare and contrast the use of treatment intervention removals with the use of comparison groups in enhancing the reliability of quasi-experimental research designs.

3. What are the strengths and weaknesses of a regression-discontinuity design relative to a randomized experiment?
4. Why might a researcher decide to choose a quasi-experimental design over a random experimental design?

Discussion Questions

1. Why are quasi experiments sometimes more appropriate than randomized experiments?
2. Why are interrupted time-series designs believed to be inferentially stronger than before-after research designs? What is a disadvantage of an ITS design relative to before-after designs?
3. How do cohorts mitigate internal validity problems associated with nonequivalent group designs?

Exercises

1. A major foundation has established a matching grant program for children in the Head Start program. The parents of children in six preschool centers will receive \$2.00 from the foundation for every \$1.00 they place in a special educational savings fund for their child. The money in the funds will become available to the children when they reach the age of 18. The foundation would like you to design an evaluation of how well the matching program works in the six centers over a five-year period. Pick a quasi-experimental design that you would like to use and prepare a presentation to the foundation directors justifying your choice.
2. A member of Congress who chairs the congressional committee that funds Medicare is interested in the effectiveness of a new weight control medication that the Medicare program has just begun covering. Of particular interest is the change in the health of recipients of the new drug after six months and after twelve months. The Center for Medicare and Medicaid Services has retained you to conduct an evaluation of the effectiveness of the medication when given to Medicare patients and prepare a report that can be used to report back to Congress. Identify some of the threats to internal validity that might arise from the design of such an evaluation and how you might overcome these threats.

Contributors: John Kennedy and Nancy A. Potok

CHAPTER 6: USING RANDOMIZED EXPERIMENTS

Learning Objectives

This chapter will help students gain an understanding of the following:

- Under what conditions experiments are useful
- Strengths and weaknesses of using randomized experiments as a research tool
- Possible solutions to challenges inherent in randomized experiments

Key Terms

- Program impact
- Randomized experiment
- Random assignment
- Spillover effects
- Nonparticipation
- Crossovers
- Attrition
- Matched pairs

Review Questions

1. What are the benefits of randomized experimental studies?
2. In what ways can spillover effects harm an experiment? Can they be avoided?
3. What are some of the ways that bias can be introduced during the implementation of randomized designs? When does sample attrition not necessarily bias results?
4. What are the weaknesses of randomized experiments?
5. How might one ensure the integrity of the random assignment process?
6. What should be included in a written evaluation agreement?

Discussion Questions

1. Why are experiments useful in measuring impact only? Why are they not helpful in measuring customer satisfaction, opinions, or measuring the inner strengths of a program?
2. Is it ethical to randomly select an experiment population for a treatment from which they greatly benefit, when others who want to participate in the study and are in need of treatment are turned away? Under what circumstances can you foresee this happening?
3. In what ways can stakeholder buy-in surrounding a randomized experiment be promoted?

Exercises

1. A Florida school district wants to know if a service-learning program would successfully engage otherwise uninvolved youth in order to improve their attitude toward community service. Depending on the results, they will implement a mandatory community service requirement for graduating seniors. They are currently planning on letting students voluntarily sign up for service learning. Draft a proposal to the school administrator explaining why a randomized experiment evaluation is the most appropriate tool, as opposed to simply letting teens sign up for the program, addressing validity issues, costs, and what form of randomization would be most appropriate to meet their needs.
2. Make a chart of the potential threats to internal and external validity that are common when using randomized experiments. Identify specific scenarios in which these would be likely to happen. Next, being as specific as possible, write a limitations explanation (as would be seen in the Limitations section of an analysis) on how each threat specifically affects the legitimacy of the data.

Contributor: Rachel Wyatt

CHAPTER 7: META-ANALYSIS, SYSTEMATIC REVIEWS, AND RESEARCH SYNTHESSES

Learning Objectives

This chapter will help students gain an understanding of the following:

- **Difference between meta-analysis, systematic reviews, and research syntheses**
- **Importance of being conscientious in reviews**
- **Resources that can be employed to conduct a thorough review**

Key Terms

- **Systematic review**
- **Meta-analysis**
- **Research synthesis**
- **Bias**
- **Management strategy**
- **Analysis strategy**
- **Model programs**
- **Exemplary programs**

Review Questions

1. List examples of "dry land," "deep water," and "swamp" with reference to presence of evidence as defined in this chapter; how can one tell the differences between the three types of evidence?
2. List the different parts of a meta-analysis, systematic review, or research synthesis that one can help contribute to in a study. What might one be able to learn as a researcher from participating in the production of each section?
3. What is the relationship between exemplary or model interventions and systematic reviews, meta-analyses, and syntheses?

4. What are some by-products of organized efforts to produce systematic reviews and syntheses? After naming the by-products mentioned in the chapter, brainstorm to list by-products not mentioned in the chapter.

Discussion Questions

1. What are some reasons that contributing to a meta-analysis, systematic review, or research synthesis may be a better way of learning about such techniques than reading or taking a course alone?
2. Discuss the some of the flaws that can be found in conventional literature reviews. What steps can be taken to ensure such flaws are minimized?
3. What can be learned from a disciplined meta-analysis or systematic review?

Exercises

1. Minimum sentencing laws for certain crimes are in place in many states. Having observed that such minimum sentencing laws seem to lower crime rates in some states, the Florida House of Delegates has commissioned you to conduct a research synthesis evaluating minimum sentencing laws. What is your responsibility as a researcher in conducting this research synthesis? To what extent are you responsible for the validity and reliability of the studies you include in your analysis? What authority do you have as a researcher in not including previously conducted research in your research synthesis? Suppose that when you are commissioned for this research synthesis it is implied that you are expected to find that minimum sentencing laws do, in fact, reduce crime rates. Does the authority you have as a researcher change if you must answer to state representatives who have commissioned you for this analysis?
2. The Environmental Protection Agency (EPA) is researching the quality of air in urban areas and the impact of public transportation on the quality of air within these areas. It has asked you to conduct a systematic review of the air quality effects of public transportation in cities with populations of over one million. Recognizing that the EPA is a national agency within the United States, what are the advantages and disadvantages of including international resources and studies with national resources and studies? You have decided to include international studies dealing with international cities within your systematic review; to what extent must you pay attention to the different language and terminology used in different analyses and

reviews to ensure validity and reliability in your systematic review? To what extent must you pay attention if you only included national reviews? To what extent does validity and reliability in a single study impact a meta-analysis, systematic review, or research synthesis overall? Must one take into account international resources and studies if one is focused at the state or local level, and vice versa? If so, how?

Contributor: Catarina Bummara

CHAPTER 8: TRAINED OBSERVER RATINGS

Learning Objectives

This chapter will help students gain an understand of the following:

- Applications for using trained observers
- Advantages, disadvantages, and limitations of trained observer ratings
- Proper design and implementation of trained observer ratings

Key Terms

- Trained observer procedures
- Trained observer ratings
- Inter-rater comparability
- Complex scales
- Anchoring scales
- Simple scales

Review Questions

1. Identify skills needed by trained observers and explain why these skills are needed.
2. Describe rating scales generally used by trained observers and explain why the scales are designed as you have described.
3. What types of sources might be helpful to identify the most relevant attributes and conditions that should be reflected in ratings?
4. Compare and contrast single-attribute with multi-attribute rating scales.

Discussion Questions

1. Table 8.1 in the book lists some characteristics rated by trained observers. Name three more characteristics to which trained observer ratings could be applied. Name three characteristics to which one would not want to apply trained observer ratings and explain why.
2. Discuss the strengths and weaknesses of various methods for aggregating the ratings of individual raters into an overall assessment.

Exercises

1. The department of social services in a major city provides grants to homeless shelters that are operated by a variety of nonprofit organizations. The grants are used to cover the operating costs of providing meals, sleeping accommodations, and bathroom facilities for families in need of transitional housing. The city has detailed standards on the level of services provided by the shelters, and has asked you to conduct an evaluation of how well the shelters are meeting those standards. Describe how you would set up an evaluation using trained observers, including the type of rating system you would use.
2. The city manager of a medium-sized community instituted a major street improvement program six months ago in response to citizen complaints about potholes, abandoned vehicles, late garbage pickups, and litter. The city manager would now like to assess how well the initiatives are working. She has asked you to conduct an evaluation of the street improvement program. Create a worksheet that can be used by trained observers to rate the street improvements.

Contributor: Nancy A. Potok

CHAPTER 9: USING SURVEYS

Learning Objectives

This chapter will help students understand the following:

- How to determine when it is appropriate to use a survey
- Useful steps for survey construction
- Types of surveys and when each should be applied
- When contracting out a survey is most effective

Key Terms

- Survey reliability
- Survey validity
- Simple random sample
- Stratified random sample
- Multistage sample design
- Convenience sample
- Quality Control
- Response Rates

Review Questions

1. What are some common mistakes in writing survey questions? What are some strategies avoid them?
2. What is quality control in the survey context and how is it accomplished?
3. Why are response rates important, and what are some methods that can be employed to try to increase them?
4. Explain the logistical issues that are involved in obtaining a high response rate.

Discussion Questions

1. Have you ever been surveyed? If you have, what was your experience like and how do you think the survey takers conformed to your notions of a valid survey?
2. What are some incentives you like to receive for doing a survey?
3. How can a survey corroborate evaluation results?

Exercises

1. The National Institutes of Health, a government agency, has a Small Business Office that provides a seminar each month on how small businesses can obtain government contracts. Each month about 45-150 small business owners attend. The Small Business Office is looking for ways to improve the seminar and has sought your survey expertise to create a survey as an evaluation tool. The Small Business office keeps a file of contact information of every person who has attended seminars in the past five years. Explain what steps you will take to create the survey. Include examples of survey questions; identify the population; explain how you will obtain a proper sample; and indicate how you will insure the validity and reliability of the data gathered.
2. Legal Services for D.C. Prisoners is a nonprofit organization that advocates on behalf of prisoners who lived in Washington, D.C. Recently the Federal Bureau of Prisons adopted a policy to place D.C. prisoners in federal prison facilities located throughout the country. In an effort to evaluate how the nonprofit can advocate on behalf of prisoners and their families affected by this policy, you have been hired to create a survey. The survey will be distributed to a sample of families in the D.C. area. The goal of the survey is to find out what the main concerns of the family members are and how the nonprofit can adjust its organization to better carry out its advocacy mission. Explain what steps you will take to create and administer the survey. Include examples of survey questions, identify the population from which you will sample, explain how you will obtain a proper sample, and describe how you will ensure the validity and reliability of the data gathered.

Contributor: Patrick Wells

CHAPTER 10: USING EXPERT JUDGEMENT

Learning Objectives

This chapter will help students understand the following:

- **When to use an expert evaluation**
- **Tradeoffs between validity, reliability, time required, costs, and clarity when using expert judgment as an evaluation method**
- **Various procedures for using technical experts**
- **Desirable characteristics of experts**

Key Terms

- **Program uncertainty conditions**
- **Outside technical experts**
- **Program Assessment Rating Tool (PART)**
- **Informed dialogue**
- **Mertonian norms of disinterestedness**
- **Unstructured, direct interaction**
- **Structured, indirect interactions**
- **Delphi procedure**
- **Devil's advocate**
- **Dialectical inquiry**
- **Coherence**
- **Reliability**
- **Validity**

Review Questions

1. What are the characteristics of a program that would lead it to be characterized as having a high level of uncertainty regarding its effectiveness?

2. What are some methods for assessing the validity of evaluations conducted by an agency's internal program evaluation staff?
3. How might an agency go about selecting outside experts for purposes of evaluating its programs?
4. Compare and contrast the characteristics of an unstructured, direct interaction and a structured, indirect interaction.
5. When would it be beneficial to use an evaluation procedure with built-in conflict?
6. Name the three traits and characteristics of experts that are desirable for expert program evaluation.

Discussion Questions

1. Discuss three specific situations where you might want to bring in outside experts. How would you structure getting their input in each of these evaluations?
2. What are some of the benefits and constraints of using a technique such as Delphi? In what types of situations might you want a more unstructured interaction between experts in which participants must provide their reasoning and assumptions?
3. How might you go about testing the validity of expert judgments?

Exercises

1. The U. S. Census Bureau has recently fielded a new annual national survey that collects a wide variety of demographic information used by other federal agencies to allocate grants to state and local governments. The Congress, which appropriates the money for the various grant programs, has asked the Census Bureau to provide an evaluation of the new survey using outside experts as well as input from its internal program evaluation staff. In turn, the Census Bureau has hired you to design and conduct the evaluation. Design the component of the evaluation using outside experts, describing how you would select the experts and what sort of procedures you would use to elicit their judgment on the new survey.
2. The Oakland County School District has set up a series of alternative schools for students who, for various reasons, cannot function in a regular school setting.

The state board of education has asked the county to conduct an evaluation on the student outcomes for the population of students attending these schools. Because the curriculum at the alternative schools differs somewhat from the curriculum at the other schools in the district, the school superintendent would like to supplement standardized achievement tests with an evaluation of the program by outside experts. Select the members of the panel and create the list of topics you would like the panel to cover. Because you have decided to bring the panel together for face-to-face discussions, provide the initial list of questions that the panel will be considering.

Contributor: Nancy A. Potok

CHAPTER 11: ROLE PLAYING

Learning Objectives

This chapter will help students understand the following:

- Identify what role playing is
- Proper application of role playing
- How to select role players
- When to use role players for evaluation
- When role playing can be coupled with statistical analysis

Key Terms

- Paired role playing
- Sampling

Review Questions

1. How is role playing an evaluation tool?
2. What kinds of characteristics matter when choosing a role player and why are they important?
3. How does sampling affect role playing?

Discussion Questions

1. What ethical considerations should you be aware of when role playing?
2. What are some reasons for employing role playing?
3. After role players have been systematically trained, why would it be important to test their reliability?

Exercises

1. The Centers for Medicare and Medicaid Services (CMS) has created a new help line to assist elderly citizens with questions they may have about the new prescription drug benefit cards. CMS has found that it is cheapest to use non-government contractors to run the help line. Recently CMS has received some complaints about poor information being given out systematically to elderly as opposed to younger people who call on behalf of their elderly parents. In order to assess whether the help line is giving out correct information to elderly callers, you have been hired to evaluate the help line responses. You are a private consultant who specializes in using role players for evaluation purposes to find out if there truly is a bias or if something else is going on. First, explain why role playing is the proper evaluation technique. Then, explain what steps you will take, including how you will obtain role players, what characteristics they should have, and how you will insure the reliability of the data gathered by the role players.
2. A car dealership in major metropolitan area has allegedly been discriminating against minority car purchasers. The allegations stem from complaints that minority car purchasers have been given higher interest rates for their vehicles than nonminority car purchasers. You have been hired as a private consultant by the Better Business Bureau to find out if the allegations are true. Your expertise is in using role players in such situations. Explain what steps you will take, including how you will obtain role players, what characteristics they should have, and how you will insure the reliability of the data gathered by the role players. Include why you feel that role playing is or is not the proper evaluation technique.

Contributor: Patrick Wells

CHAPTER 12: USING FOCUS GROUPS

Learning Objectives

This chapter will help students gain an understanding of the following:

- The key principles of focus group design
- When and why focus groups are used
- How to analyze and use focus group data

Key Terms

- Group dynamics
- Exploratory focus groups
- Moderator
- Probing techniques
- Reinforcing techniques
- Nonverbal communication
- Screening questionnaire

Review Questions

1. What advantages do focus groups have over other forms of data collection, such as surveys and interviews?
2. Why is it important for evaluators to be concerned with group dynamics when conducting focus groups?
3. What are important considerations for evaluators to keep in mind when formulating focus group questions?
4. What considerations should be made when deciding whether to use a contractor to conduct a focus group?
5. When deciding on the composition of a focus group, should evaluators look for homogeneity or heterogeneity among the participants? Why?

Discussion Questions

1. For what sorts of evaluations are focus groups the best method of data collection?
2. In what ways can the participants' backgrounds affect the dynamic of a focus group?
3. How can the evaluator involve key stakeholders in the selection of the focus group participants?
4. How might a moderator handle a situation in which one group member offends another? How might this scenario affect the data collected?
5. In what ways can the moderator get participants to contribute more to the conversation?

Exercises

1. You are conducting an evaluation for the government of the District of Columbia of a program to attract more middle class families to live in the city. After analyzing your quantitative data you have found that the program is not achieving its goals. In order to learn more about why that is so, you have decided to conduct focus groups. However, to save money, your client would like you to send out a mail survey to get the answers you're looking for instead. Write a one-page letter explaining why you feel it would be worth the extra funds to conduct the focus groups.
2. The U.S. Department of Agriculture's Conservation Reserve Program allows farmers to set aside eligible portions of their land for conservation purposes, such as creating buffers for existing streams and rivers, in exchange for payment. Recently, an Internet-based automated system was developed to make it easier for farmers to enroll their land in the program. By logging onto the Web site, farmers can type in the longitude and latitude of their land and immediately learn whether their land is eligible, and what the rental rate would be if it were enrolled. The USDA conducted an internal evaluation to determine whether enrollment increased after the system was established. They found that enrollment went up significantly in some areas, but very little or not at all in others. You are the contractor hired by the USDA to find out why farmers in certain areas are not using the system. You have decided to use focus groups. Explain what steps you will take, including how you will locate your participants, where in the community you will hold the focus groups, and what your questions will be.

Contributor: Linda Miller

Chapter 13: Collecting Data in the Field

Learning Objectives

This chapter will help students gain an understanding of the following:

- Relationship between data collection objectives and fieldwork procedures
- Weighing cost and time constraints when designing fieldwork procedures and objectives
- Importance of attention to procedural and logistical details when planning, executing, and following-up on fieldwork
- How to ensure consistency and reliability in data collection and analysis

Key Terms

- Data collection
- Qualitative data
- Quantitative data
- Fieldwork
- Research question
- Unit of analysis
- Site
- Sample
- Semistructured instruments
- Structured instruments
- Respondent
- Confidentiality
- Daily review coding

Review Questions

1. What are two types of fieldwork commonly conducted and in what ways are they similar? In what ways do they differ? Why?
2. In what ways do cost and time constraints affect choices made in designing data collection in the field?

3. Why are strict procedures important in the collection of data in the field?
4. How can evaluators ensure that collected data is reliably and consistently analyzed?

Discussion Questions

1. Discuss the implications for data analysis in the use of both semistructured and structured data collection instruments.
2. What are the ethical implications of sharing preliminary findings with program staff before analysis has been conducted?
3. Participatory research methodologies have become increasingly widespread in fieldwork and data collection. What are the positive and negative aspects of including program staff and other stakeholders in the data collection effort?

Exercises

1. You are a member of research team hired by a local nonprofit social welfare organization to evaluate its performance working with illegal immigrants and transient laborers from Latin America in and around San Diego, California. The organization helps its clients obtain health services and housing, as well as fulfill their daily basic needs. Specifically, the evaluation seeks to determine whether it is meeting the needs of its beneficiaries and targeting the correct population. The organization would also like to understand how it is perceived among the beneficiary population and the level of trust it has gained among members of this highly fluid and wary population. Design two interview guides, one for field staff members and another for clients. What are some of the most important issues to consider when conducting these interviews? In what ways do they differ?
2. For the above scenario, write two statements of confidentiality for each response group. Be sure to address issues specific to the separate groups in the separate statements of confidentiality.

Contributor: Benjamin B. Friedman

CHAPTER 14: USING AGENCY RECORDS

Learning Objectives

This chapter will help students gain an understanding of the following:

- Evaluator reliance on agency records and their detail, availability, and accuracy
- A disciplined process of scrutinizing data parameters, definitions, and sources
- Teamwork and building effective win-win strategies with agencies in order to obtain complete and accurate information
- Confidentiality issues, and strategies to mitigate them
- Techniques to ensure accurate data transposition

Key Terms

- Aggregated data
- Weighting
- Duplicate counts
- Data checks/ cleaning data

Review Questions

1. In what ways are data comparisons affected by sources originating in and outside an agency?
2. Why can issues of confidentiality complicate data collection procedures? What can be done to mitigate these problems?
3. Explain the importance of concrete definitions when evaluating data collected over time or across program boundaries or jurisdictions.
4. Why are relationships between the evaluator, evaluation team, and client crucial to effective data collection and interpretation?

Discussion Questions

1. How can evaluators handle missing or incomplete archival information?
2. What special problems are posed by "aggregated" data? What are some alternative solutions to dealing with problems of aggregation?
3. What strategies can be employed to protect client privacy?
4. Discuss the concepts of value ranges, "reasonableness," and data consistency as they relate to cleaning the data.

Exercises

1. A group of women in a federal agency have threatened to file a class action suit against the agency, claiming that women are discriminated against in promotions and awards. Before filing suit, however, they have agreed to view the outcome of an evaluation that the agency will have conducted to assess the promotion and award outcomes in the agency for bias. The evaluation is to be conducted by a neutral outside evaluator—you. The evaluation will use the agency's records as the primary source of data for the evaluation. Describe some of the challenges you will face in conducting this evaluation and how you plan to overcome those challenges. Present your approach in a presentation that you will give to both the agency management and the employee group.
2. A new federal law is being considered that would mandate that fathers be involved in some way with their children when the children have been placed in foster care. Some members of Congress, buttressed by groups supporting children's rights, are strongly against this law, contending fathers are often a negative influence on the children. Other members of Congress, buttressed by groups representing noncustodial fathers, are supporting the law, contending that fathers are a strong positive presence for their children, particularly if they've been placed in foster care due to action on the part of the custodial parent, most often the mother. You have been asked by the Department of Health and Human Services to conduct an evaluation using records in cooperating states to assess whether the influence of fathers has been positive or negative. Explain how you plan to go about conducting this evaluation and what the potential risks are. Include how you plan to try to mitigate some of the risks where possible.

Contributors: John Kennedy and Nancy A. Potok

CHAPTER 15: QUALITATIVE DATA ANALYSIS

Learning Objectives

This chapter will help students gain an understanding of the following:

- Key pre-analysis elements for producing good qualitative data analysis
- How to organize and code data to facilitate the process of analysis
- How to produce objective analytic conclusions and communicate them effectively
- The various threats to objectivity and validity and means of countering those threats
- Methods to ensure effective presentation of a final analytic product

Key Terms

- Qualitative analysis
- Coding
- Data displays
- Threats to objectivity
- Threats to validity
- Internal validity
- External validity
- Validity
- Vignette
- Milestones

Review Questions

1. What are the four pre-analysis elements that affect the quality of analysis in a program evaluation and why are they important?
2. How is the production and communication of objective analytic conclusions aided by data reduction and pattern identification before and during qualitative data analysis?
3. Why are objectivity and validity important factors to consider in data analysis?

4. In presenting the final findings, what are some of the key strategies to ensuring that the production process is smooth and efficient?

Discussion Questions

1. How do the four key pre-analysis elements assist in the compilation and presentation of final research findings?
2. Explain why objectivity and validity are important in data analysis. Discuss the effects that different threats can have on research and ways to counter those threats.
3. How would you include stakeholders or participants in a qualitative data analysis process? Is their role only as informants or can participants also conduct valid analysis of relevant data? What is their relationship with researchers then?

Exercises

1. Identify possible coding concepts in the following interview excerpt. Be sure to make it clear from where in the text the concept is taken, either by duplicating the structure shown in Exhibit 15.1 or through some other means (line reference, and so on).

"Ms. Jones said that when she started work about a year ago, she was surprised at the disorganization and lack of guidance from program directors in terms of strategy and long-term goals for prevention programs at the youth center. From what she understood, the mission statement and strategy of the organization have not been revisited or modified in a number of years. Primary goals included educating youth on pregnancy and STD prevention as well as reducing risk factors through life-skills education. Despite the plethora of educational materials that were in the storage room, no related educational and prevention activities were being held or had been held in the past year. Instead the youth center had become a recreational center that provided youth with a place to play, but little else. Actually computer classes were held, but on Macintosh computers that were old and didn't even have current software. Other employees were also lackadaisical and didn't appear to take their jobs seriously. When she shared her concerns with her coworkers she was told, 'That's how we do things around here.' When she expressed these concerns to her supervisors at the organization headquarters, she thought that they were attentive and appeared to understand what she had said, but nothing had changed in the intervening six months. One concern that Ms. Jones had was that outcomes could not be measured and thus further funding could not be secured. She thinks that the program and the youth center is a good idea, but unless activities are aligned with program goals, she will have to start looking for another job."

2. Your six-person team of crack evaluation specialists was contracted by an international nongovernmental organization to conduct an implementation evaluation of a neo-natal health program in Nicaragua, a Central American country suffering from high rates of poverty and malnutrition. Research was conducted in six sites in different parts of the country: two rural, two urban, and two in large towns in rural areas. You had a six-month timeline and \$60,000. Semistructured interviews were conducted with program staff and surveys were administered to program participants in all sites. You also collected demographic and economic data about the various communities in which you worked. Based on the guidelines set forth in this chapter, describe the analysis process you would have gone through in this situation and the products you might have produced. Discuss the benefits and costs of the different practices summarized in Table 15.3.

Contributor: Benjamin B. Friedman

CHAPTER 16: USING STATISTICS IN EVALUATION

Learning Objectives

This chapter will help students gain an understanding of the following:

- How to match analytical techniques to the level of measurement, audience, and evaluation questions
- Background for making statistical decisions such as determining the right sample size, selecting a statistical confidence level, and testing statistical significance
- Four principles for selecting samples
- How to appropriately report statistics

Key Terms

- Levels of measurement
- Descriptive statistics
- Inferential statistics
- Probability sample
- Stratified sampling

Review Questions

1. What are the four levels of measurement identified by Stevens? Which are categorical and which are continuous? Give an example of data that might fit into each category.
2. Explain the four principles that should guide evaluators when they select samples.
3. When is it appropriate to use a chi-square test and what are the assumptions associated with using that test?
4. Describe the data analysis techniques most likely to be used with ordinal measures and with nominal measures.

Discussion Questions

1. Table 16.1 in the text lists design features likely to generate false positives or false negatives regarding tests of the null hypothesis. For each of the design features, give some examples of studies where such features might be incorporated and think of some ways to mitigate the risks. For example, you are conducting an evaluation of services being provided by a clinic by observing clinic operations. How would you try to avoid the Hawthorne effect?
2. What are the pros and cons of using various confidence levels when testing hypotheses? When might an 80 percent confidence level be sufficient rather than a 95 percent confidence level?
3. Try to determine what appropriate sample sizes might be for various types of analyses. For example, if you wanted to evaluate a technique for teaching reading to first graders used in a school district with twelve hundred first graders, what would be an appropriate sample size of first graders to include in your study? If you wanted to evaluate school voucher programs nationally, what would your sample include?

Exercises

1. A national association of lawyers has been running an educational program and support group for women and minority attorneys who work for large law firms to help them become partners in these firms or otherwise advance their careers. The association believes that many large firms have a glass ceiling for these employees, and would like to find out if its educational and support program is making a difference. They have asked you to evaluate the effectiveness of the program by collecting information from a sample of their members—including males and females, non-minorities and minorities, participants and nonparticipants—to compare career outcomes. Before you begin, they have asked you to speak to the association's board of directors and explain what you will be doing.

Prepare the presentation you would give to the board of directors. Be sure to include important information such as that recommended in the tips for presenting data analyses in Chapter 16, as well as your design and methodology for the evaluation.

2. The Department of Homeland Security has asked you to prepare a proposal to conduct an evaluation of its grant program to local governments to purchase equipment for emergency first responders, such as police and fire departments. The Department has given out grants totaling \$200 million to thirty thousand municipalities over the last three years. It would like to know if local

governments have used the grants to purchase equipment, whether the equipment fits the Department's eligibility criteria, and whether the municipalities are using the equipment appropriately. There is not enough money in your contract to contact all 30,000 municipalities. Prepare a proposal to the Department that describes what hypotheses you would be testing, how you would size and design a sample for this evaluation, what confidence interval would be acceptable, and what analytical techniques you would use on the data you collect.

Contributor: Nancy A. Potok

Supplement to the *Handbook of Practical Program Evaluation*, Second Edition, edited by Joseph S. Wholey, Harry P. Hatry, and Kathryn E. Newcomer, ISBN 0-7879-6713-0, cloth, © 2004 Jossey-Bass, An Imprint of Wiley.

CHAPTER 17: USING REGRESSION ANALYSIS

Learning Objectives

This chapter will help students gain an understanding of the following:

- When it is appropriate to use regression analysis for change statistics
- Basic syntax for using SPSS to conduct regression analysis
- Testing a causal model on the effects of a program
- Issues that can affect the accuracy of a regression analysis

Key Terms

- Multiple regression
- Standardized residuals
- Slope
- Intercept
- Regression coefficient
- Standard error
- Mediation
- Correlation
- Causation
- Outliers
- Dummy variables
- Predictor variables
- Dependent and independent variables
- Power analysis

Review Questions

1. Compare and contrast the characteristics of an independent samples t test with a multiple regression analysis. In what types of analyses would you use each of these?

2. Describe the characteristics of the coefficient of determination and what information it provides.
3. What important information is provided by a mediation analysis?
4. What is meant by the power of a statistical test? Name some methods that one might use to overcome the problem of low power.
5. Describe some ways that data collected for a study might interact that could cause you to inaccurately describe whether one variable is a good predictor of a particular outcome.

Discussion Questions

1. You are asked to evaluate the effectiveness of a job training program offered to nonviolent offenders during the last six months of their incarceration. The basic design of your study calls for you to track the prisoners during the first three months after their release to determine whether they were able to find employment in the field for which they were trained. How might you determine whether the job training program was a mediating factor in ex-offenders' ability to find post-release employment?
2. You are evaluating whether the running heart rate of men between the ages of forty-five and sixty-five is a good predictor of blood pressure. The men in the study are put on a treadmill and the following data are collected: resting heart rate before exercise, heart rate after five minutes of running, blood pressure before and during the test, age, height, weight, profession, marital status, parental status, race, and ethnicity. Discuss areas where you might expect to find multicollinearity.
3. Regression analyses are based on several assumptions about the data, for example, that there is a linear relationship between the dependent and the independent variables. Identify the key assumptions and discuss instances where the assumptions might not hold.

Exercises

1. You have just completed a study of a sample of current college sophomores attending four University of Maryland-system campuses, to determine how well their SAT scores correspond to their grades during their first 3 semesters at college. Your data show that 50 percent of students who had scored 1,000 or higher maintained a B average for the semesters included in the study, 75 percent of students who had scored 1,150 or higher on their SATs maintained a

B average, and 85% of students who scored over 1,300 maintained a B average. About 30 percent of the students in the study participated in special sessions for freshmen designed to help them acclimate to campus and manage their workload.

What tests would you run on the data to determine whether SAT scores are a good predictor of college performance? Describe the tests and the results you might be looking for.

2. A family planning organization funded by a large foundation is interested in determining whether the number of children in a family is a good predictor of whether the children will graduate from college. You have been asked to analyze data from a recent national survey of college graduates that collected the number of siblings of each respondent. The survey also collected other demographic characteristics of the respondents such as age, sex, race, and educational attainment of parents and older siblings.

Design a regression analysis to make this determination. If you have knowledge of SPSS or SAS, try to write a program that would run the tests of the data that you choose to include. Discuss what types of results you would be looking for and how you would look for irregularities in the results.

Contributor: Nancy A. Potok

CHAPTER 18: COST EFFECTIVENESS AND COST-BENEFIT ANALYSIS

Learning Objectives

This chapter will help students gain an understanding of the following:

- Whether cost-effectiveness analysis or cost-benefit analysis is most appropriate when evaluating program effectiveness and economy
- Application of various economic concepts, such as Net Present Value, sunk cost, discounting, and opportunity cost
- Defining all direct and indirect costs associated with program objectives
- Limitations of quantifying costs and benefits and the strategies that can be employed to mitigate these limitations

Key Terms

- Cost-benefit ratio
- Cost-effectiveness ratio
- Discount rate
- Marginal cost
- Marginal benefit
- Market value
- Net present value
- Opportunity cost
- Sensitivity analysis
- Shadow pricing
- Sunk cost

Review Questions

1. What is the difference between “real” benefits and costs versus “transfers?”

2. Why is it important to capture externalities such as spillovers when determining costs and benefits?
3. Why do multiple benefits make the use of cost effectiveness analyses problematic?
4. What is depreciation, and how is it assigned as a program cost?
5. How is the cost-benefit ratio calculated? How is a cost-benefit ratio useful to evaluators?
6. How do issues of equity complicate applying the results of cost-benefit analyses?
7. What is "opportunity cost," and how is it related to Net Present Value (NPV)?
9. What is "market value," and why is it normally used in the valuation of a benefit?
10. Why is caution advised when assigning market values to benefits?

Discussion Questions

1. As an evaluator conducting a cost-effectiveness or cost-benefit analysis, in what ways can you ensure decision makers have the relevant information from which your analysis is based on?
2. In what circumstances is cost-effectiveness a more useful analysis than cost-benefit analysis, and why is there a need for both?
3. In what ways can estimation of various program costs be broken down and why is it important to separate these costs?
4. Why is it typically more difficult to measure benefits in government than in the private sector? Explain the various types of benefits involved.

Exercises

1. The city of Branson, Missouri, is considering building a major multi-use convention center to attract more tourists, events, and conferences. With over sixty live shows, Branson is considered "The Live Music Show Capital of the World." Two different firms have made proposals to construct the facility.
Project 1: The Ozark Construction and Development Company proposes to construct a convention center that will cost \$20 million to build (borrowed at a 5 percent interest rate) and will have operating costs of \$3 million annually. It is

projected to generate three benefits: fees for use (\$2 million annually); an increase in tourism economy (\$4 million annually); and an increase in property values in the surrounding area (\$1 million annually).

Project 2: The Sheppard of the Hills Construction and Development Company plans to build a more elaborate convention center that will include two attached hotels. This facility will cost \$35 million to build (also borrowed at 5 percent interest rate) and will have operating costs of \$5.5 million annually. It is projected to generate three benefits as well: fees for use and hotel rooms (\$7 million annually); an increase in tourism economy (\$4 million annually); and an increase in property values in the surrounding area (\$1.2 million annually).

You have been asked by Mayor Andy Williams to evaluate the two proposals over a 5-year period using cost-benefit analysis. You should include assumptions made and use a 5 percent discount rate.

2. As a volunteer for Waldorf Community Church, you have been asked to organize the church's annual summer Vacation Bible School (VBS). The goal of VBS is to provide Biblical learning opportunities to children of church families as well as neighborhood children who do not normally attend the church. Historically, about seventy-five children are ministered to during VBS and the church board typically budgets \$1500 in current dollar value. Of the seventy-five children who usually attend VBS, fifty are typically from Waldorf Community Church families, ten are children from other churches, and fifteen are from families not currently attending a church.

Two curricula appear to have the most potential for meeting VBS goals within budget:

Title	Cost of Materials	Staff	Facility Size	Max Students
"Streets of Gold"	\$1500	12	2000 sq. ft.	75
"This Little Light of Mine"	\$1500	16	2500 sq. ft.	90

The church has a 2100 square foot area where VBS is traditionally held, which can accommodate eighty children and fourteen volunteers. The church also has an outdoor tent that could expand the usable space available to VBS to 2500 square feet. The tent, however, is often used by the church youth group to host car wash fundraisers.

Despite the space limitations, higher curriculum cost, and requirement for more staff, you and your volunteers prefer the "This Little Light of Mine" curriculum for this year's VBS. Initial estimates show that up to eighty-five children may attend

this year's VBS, although the lack of an RSVP requirement makes precise forecasting difficult.

Historically, a core group of 13 trained volunteers have been available to help administer the church's VBS program. However, four untrained volunteers have expressed an interest in assisting with VBS if the size of the program requires additional volunteers. It takes approximately four hours to train new VBS volunteers. Training would take place on a Saturday morning.

Paul Charles, the church pastor, is interested in growing the church's VBS program and wants to support the "This Little Light of Mine" curriculum. However, he has asked you to prepare a cost-effectiveness analysis to present to the church board, justifying this selection.

In your presentation, consider the direct and indirect costs and benefits associated with the increased facility requirements, additional staff (and training); the fixed and variable costs corresponding with numbers of students who may attend (that is, the marginal benefit-cost per additional student); and opportunity costs. In your presentation, be sure to discuss the impact of spillover costs associated with students from other local churches attending Waldorf Community Church's VBS, and ways these costs could be mitigated.

Contributors: John Almond and John Kennedy

CHAPTER 19: PITFALLS OF EVALUATION

Learning Objectives

This chapter will help students gain an understanding of the following:

- **Hallmarks of methodological integrity**
- **Typical limitations to validity and reliability of data findings occurring prior to, during, and after data collection**

Key Terms

- **Credibility**
- **External validity**
- **Internal validity**
- **Measurement validity**
- **Reliability**
- **Replicability**
- **Statistical conclusion validity**

Review Questions

1. Why is it essential to anticipate who the end users of an evaluation are likely to be during evaluation design?
2. What is the importance of defining an accurate scope of work (terms of reference)?

Discussion Questions

1. Why is generalizing the results of an evaluation frequently problematic and how might you enhance the likelihood of generalizability of results?
2. Discuss how you would use the advice provided to train your evaluation team prior to beginning data collection.

Exercise

1. You are thinking of bidding for an evaluation job in response to an ad to conduct an evaluation of visitor services within U.S. National Parks. The ad only specifies that the evaluation should be completed within six months from the award of contract and the funding ceiling. How would you design your methodology to take into account typical pitfalls identified in this chapter? Be realistic in terms of coverage and costs.

Contributors: Shawn Mann and Kathryn Newcomer

CHAPTER 20: MANAGING EVALUATION PROJECTS

Learning Objectives

This chapter will help students gain an understanding of the following:

- Developing rational proposals
- Clarifying the evaluation mandate
- Staffing and organizing an evaluation project
- Making assignments productive through building staff consensus
- Importance of monitoring interim progress
- Links between product quality and usefulness and involvement of stakeholders, early planning, goal alignment, and follow-up

Key Terms

- Evaluation project management
- Project director
- Evaluation staff or evaluators
- Evaluation mandate
- Sponsor
- Client
- Stakeholders
- Program

Review Questions

1. What are the five phases of proposal development and the key issues to consider during each of them?
2. Why is it important to establish and maintain agreement with the sponsor or client about the evaluation mandate?
3. What is the accepted general approach to project staffing?
4. Why are clarity, coordination, and planning essential to assignment productivity?
5. Describe common monitoring problems during evaluations and the different solutions that can be utilized in addressing those problems.

6. How do the four steps of developing an evaluation report ensure product quality and usefulness?

Discussion Questions

1. Describe the process for forming teams for evaluation research. In what way is this process essential for effective project management?
2. In what ways can project management ensure that staff skill sets are in line with evaluation project needs?
3. In what ways can evaluation managers ensure the participation and consideration of all relevant stakeholders throughout the course of an evaluation?

Exercises

1. You are an evaluation team with a project manager as your spokesperson. You are currently in the process of conducting an evaluation of Head Start programs in the District of Columbia for the D.C. Board of Education. The original mandate was to evaluate the effectiveness of the programs as applied towards improving their effectiveness and possibly expanding their reach. You had four months and \$100,000. At this stage you have begun to collect field data using surveys and informal interviews with program staff throughout the area. Unfortunately, Congress has cut the budget for Head Start programs in D.C. and expansion will be impossible. Instead, the program may have to be reduced in size. Rather than scrap the evaluation, the D.C. Board of Education would like you to revamp the evaluation and evaluate the existing program for ways in which it can be streamlined and reduce costs. How would you handle this situation? What steps could you take to adjust your evaluation so that it meets the changed needs of your client?
2. You are the Director of Operations at a private addiction rehabilitation program located in Joliet, Illinois, a working-class suburb of Chicago that has been hard hit by the loss of manufacturing jobs in recent years. Clients in this program are both court-referrals as part of probation requirements or are there in lieu of jail-time. You wish to know whether a new group treatment methodology is equally or more effective than more traditional methods. There have been no prior evaluations conducted on this methodology, but you have existing data on client recidivism, type of addiction (such as alcohol, drugs, gambling, and so on), and length of treatment. Write a request for proposal (RFP) outlining the evaluation goals and objectives, a list of expected deliverable products with a delivery schedule, and a list of expected evaluation tasks and activities. Also identify the expected evaluation design and guidelines and mechanisms for data collection. Findings must be reported to the board in six months and funding is limited to \$50,000.

Contributors: John Adam Almond & Benjamin B. Friedman

Supplement to the *Handbook of Practical Program Evaluation*, Second Edition, edited by Joseph S. Wholey, Harry P. Hatry, and Kathryn E. Newcomer, ISBN 0-7879-6713-0, cloth, © 2004 Jossey-Bass, An Imprint of Wiley.

CHAPTER 21: WRITING FOR IMPACT

Learning Objectives

This chapter will help students gain an understanding of the following:

- Three key components to effective writing
- Six basic formats to get the message of a report to its intended audience
- How to write the “killer paragraph”
- The basic graphic techniques that will help maximize reader interest

Key Terms

- The mom test
- Thought leaders
- The killer paragraph
- Report enhancements
- Power writing

Review Questions

1. Why is the “mom test” an important tool in ensuring effective writing?
2. How can an evaluator impress the thought leaders with his or her report?
3. Explain why the “killer paragraph” is so universal.
4. What is the most important document an evaluator can create and why?
5. How can an evaluator use enhancements and graphics to optimize readability?
6. What are the general guidelines of power writing and why is it used?
7. What should an evaluator do if he or she cannot limit the key findings to the recommended number of two to five?

Discussion Questions

1. Why is writing technique such a key component in evaluation reporting?
2. How might an evaluator handle a situation in which a stakeholder is doing a poor job of editing a report?
3. In what ways can unclear or long-winded writing hamper an evaluation?
4. How can an evaluator balance providing details and keeping the reader interested?
5. What would you do if your client asked you not to mention any limitations in the executive summary of a report?

Exercises

1. You have been hired to evaluate the effectiveness of a career development program for at-risk youth. The eight-week summer program was designed to address the problem of minority youth joblessness, teaching kids skills to help them find and keep jobs. Kids in the program worked seven hours per day as laborers and spent one hour each day in a career development course. The kids, who were referred to the program by school personnel, were paid for seven hours of work each day.

You conducted your data collection using surveys, interviews, and observations of the kids and instructors. The data reveal that the kids' scores between pre- and post-testing differ significantly in the areas of career planning and job search skills, which shows that the program had some effect. Further, you are pleased to find that the general job satisfaction rate is in the 70th percentile compared with the norms for laborers, indicating a moderately high degree of general job satisfaction.

Seventy-nine percent of the kids reported that they felt satisfied with their work and career development experiences and 68 percent reported feeling enhanced confidence in their job skills. The main problems the kids had with the program are related to conflicts with instructors. The results for two of the variables, career planning and job search skills, are statistically significant at the 95 percent level. However, you are somewhat concerned with threats to validity, particularly external, since there was a 19 percent attrition rate of kids throughout the program due to disciplinary problems and other summer commitments. Likewise, the two statistically significant variables may have covaried given their relatively high correlational coefficient of .82, and therefore may be measuring similar constructs.

- a. Write a mom test summary of the above evaluation results.
- b. Write a killer paragraph summary of the above evaluation results

Contributor: Linda Miller

CHAPTER 22: USING ORGANIZATIONAL REPORT CARDS

Learning Objectives

This chapter will help students gain an understanding of the following:

- Purpose of organizational report cards for both organizations as well as consumers
- Difference between organizational report cards and other performance measures
- How organizational report cards may trigger organizational change
- Factors contributing to levels of use of organizational report cards

Key Terms

- Organizational report cards
- Risk-adjusted performance measures
- Consumer report card use
- Organization report card use
- Political report card use
- Strategic use of policy information

Review Questions

1. Identify how organizational report cards differ from other performance measures.
2. Under what circumstances are consumers or purchasers most likely to use organizational report cards?
3. Under what circumstances are organizations that deliver services most likely to use organizational report cards?
4. List and define the five distinct stages of the production of an organizational report card.

5. List and define the characteristics of a good report card. In your opinion, which is the most important? Why?

Discussion Questions

1. Cite the trends that have encouraged the development of organizational report cards; in your opinion, which of these trends has encouraged the most development of organizational report card use? Support your opinion with examples from the chapter as well as examples you can think of on your own.
2. It has been cited that public policymakers, especially politicians, are the most neglected of all report card users. In what ways is it important that this group's use of organizational report cards be studied? How might this group's report card use be studied?
3. In what ways can researchers and evaluators ensure that organizational report cards are used by consumers?

Exercises

1. Imagine that when the Louisiana Department of Education publishes their next set of report cards on individual public schools, they would like to spearhead a new initiative to increase consumer use of these report cards. The department has come to you to better understand what they can do to increase parental use of these report cards. What would you explain to the Louisiana Department of Education as to why organizational report cards, as a whole, are in such limited use? Further, can a relatively small amount of organizational report card users have an impact on the quality of services that consumers receive from an organization, in this case, the public schools? How does this relate to how parents use these report cards? How can the Louisiana Department of Education find out who uses these report cards and why?
2. Suppose that the National Organization for Women (NOW) has decided to produce an organizational report card on the most women-friendly work places in the United States. As NOW is a nonprofit advocacy group, should this report card be produced by NOW? With a partner, argue that report cards should be produced by advocacy groups, with both organizations and consumers benefiting from the production of such report cards, while your partner argues that organizational report cards should not be produced by advocacy groups. If NOW does choose to produce this report cards, how should they choose their range of indicators?

Contributor: Catarina Bummara

CHAPTER 23: THE USE OF EVALUATION BY NONPROFIT ORGANIZATIONS

Learning Objectives

This chapter will help students gain an understanding of the following:

- Trends in challenges faced by nonprofits when dealing with evaluation
- Options available to nonprofits with limited resources
- Roles volunteers can play in evaluation

Key Terms

- Outcome measurement
- Volunteer

Review Questions

1. What roles can volunteers play in evaluation?
2. Why is creating a low-cost model for volunteer involvement especially important?
3. In what ways can volunteers be motivated to work on an evaluation?

Discussion Questions

1. For what reasons do nonprofits not utilize evaluations as they should?
2. What do nonprofits need to understand to recognize the importance of evaluation?
3. Should organizations be encouraged to collect data, even if the data-collection procedures are not ideal and not well maintained?

Exercises

1. You have been asked to present at a conference designed for directors of small nonprofit organizations specializing in post-conflict disaster relief. Like many nonprofits, all of these directors face the challenges of funding and overburdening staff resources. Some of the funders they have previously worked with have required nonstringent progress reports. However, the new wave of funding from USAID and the World Bank are making future funding conditional on outcome evaluations. These agencies are requiring evaluations, but are neither providing additional funding for evaluation costs, nor permitting that evaluation costs be siphoned from existing contracts. Understanding that your audience will be frustrated with this unfunded mandate and having this heavy administrative demand placed on them, what suggestions will you present to the directors to make them able to most cost-effectively fulfill the requirements of evaluation? What will you say to help them understand why this is not just a mandate but a means of having evaluations that will help their organization?
2. You are staff at Disabled American Veterans, a nonprofit that serves as advocate, free of charge, for disabled veterans with claims for benefits from the Department of Veterans Affairs. The organization trains advocates in areas such as advanced law, biology, communication, and critical writing to serve as attorney-in-fact for those veterans in need of their services. The Board of Directors wants to know if there are weaknesses in training, leading to lost claims. They have directed you to coordinate an evaluation including a review of past case failures and interviews of both trainees and customers to understand their experience. You have decided that you are going to use volunteers to help facilitate the evaluation. Make an outline as to what aspects of the evaluation volunteers could help with, what preparation or training they would need, who would provide that training, and how these things would affect costs.

CHAPTER 24: OTHER ISSUES AND TRENDS IN EVALUATION

Learning Objectives

This concluding chapter will help students gain an understanding of the following:

- Quality control of the evaluation process
- Selection and training of personnel
- Standards and ethics in evaluation work
- Incentives for evaluation
- Program evaluation versus performance monitoring
- Current and future trends in the evaluation field

Key Terms

- U.S. GAO "Yellow Book"
- Informed-consent
- Performance targets
- Retrospective evaluation
- Prospective evaluation

Review Questions

1. How does quality control of the evaluation process contribute to incentives for evaluation?
2. How can evaluation work use technologies to become more efficient?

Discussion Questions

1. Discuss how evaluation methodology might evolve in an organization you know.
2. Discuss how evaluation is useful in the context of donor assistance to developing countries.

Exercise

1. You are contracted to establish a performance monitoring system within a medium-sized local NGO in Brazil. The NGO is partnered with a larger NGO in the U.S., which has contracted you and is responsible for monitoring logging activities in a specific region of the Amazon. The U.S. NGO is interested in using the numbers of logging operations and the quantity of timber extracted to lobby the Government to regulate logging activities.

How would you recruit and train the right kind of personnel to do this job?

Crosscutting Exercises

The following exercises are more complex than those included in each separate chapter. These exercises draw upon skills and knowledge gained by mastering several chapters in the handbook. They need more time to complete and are suitable for extended class work or as assignments for group work done outside of class.

Exercise 1a

As a member of the Department of Homeland Security Working Group on Sea Port Security, you are tasked to evaluate two proposals designed to enhance the security and efficiency of cruise ships while embarking and debarking passengers. Proposal A (the status quo) involves the standard metal detector/ion scan check of persons and baggage brought aboard the vessel, coupled with a basic background check of passengers to standard law enforcement and intelligence databases. Additionally, extended searches of random passengers are conducted as a deterrent to more discreet means of secreting terrorist items aboard the vessel. The goal of Proposal A is to effectively deter terrorist events through “layered” security measures while protecting individual liberties by avoiding the profiling of individual demographic groups.

Proposal B follows the same initial baggage and personal screening as Proposal A, but does not use a secondary extended search of random passengers as a deterrent measure. Instead, Proposal B utilizes a new computer software program that combines the standard background check with a complex threat matrix that matches the traveler's previous travel record and other intelligence information with predetermined threat matrices. Based on this assessment, a “terror probability” is assigned to each passenger. Only passengers who reach a certain probability threshold will receive searches similar to the random searches employed in Proposal A. The hoped-for advantages of Proposal B are: (1) to accelerate the check-in procedure by requiring fewer unnecessary inspections that are largely symbolic and deterrent in nature; (2) to reduce the overhead cost of passenger security during check-in; and (3) to achieve more effectiveness in proactively identifying potential security problems.

Your team has been given permission to design and implement a field test to measure the effectiveness of each program. In addition to normal passenger traffic being screened, the test will include “planted” security breaches to evaluate the effectiveness of each proposal's performance to simulated security breaches. Measurements will also be taken at each passenger terminal to determine the overall speed of check-in/embarkation aboard the cruise ship in order to determine the level of efficiency of each proposal. Port(s) selected for Proposal B will have a one-week training period to acclimate security personnel to the new system prior to its full implementation.

Of the roughly two-dozen seaports embarking cruise ship passengers on the U.S. Pacific and Atlantic Coasts, each is unique in its combination of terminal layout, security service provider, and certain discretionary procedures not fully covered in the regulations. Although no two are exactly alike, terminals in Miami, Port Everglades (Ft. Lauderdale, Fla.), Port Canaveral (Cape Canaveral, Fla.), New York, and San Juan employ the same security company. Terminals in Tampa, Fla., Baltimore, Boston, Seattle, and San Diego are the most similar in terminal size and layout.

The cruise ship business can be quite seasonal, especially at the Northern ports. Typically, cruise ship traffic out of Boston, Baltimore, New York, and Seattle diminishes significantly during the winter months. As one would imagine, significant political interest surrounds the selection of ports for evaluation as well as the results. Approximately \$10 million dollars in grants from the Transportation Security Administration (TSA) are at stake, with ports selected for the study potentially having an advantage when competing for these funds in the future.

Your team has staff, funding, and time to evaluate Proposals A and B at four seaports, total. The evaluation must be completed in 12 months, and can run a maximum of six months at each selected terminal. Evaluation periods do not have to overlap. Within limitations, the study sample should be as representative as possible of a typical U.S. cruise ship terminal, in order to generalize results throughout the industry with as much confidence as possible.

Construct interrupted time-series designs using four of the seaports, including an appropriate schematic (for example, XO, with X being the treatment). Construct one design without comparison groups, and one design with comparison groups. Discuss the strengths and weaknesses of the use of each design, and propose mitigating steps (such as multiple interventions) that might be used to improve design performance. Discuss threats to validity associated with these methodologies, including pre- and post-testing. Be sure to factor in issues of political feasibility and the generalizability of study results to the entire industry.

Exercise 1b

Change the scenario in Exercise 1a so that during initial beta testing of your designs, the security companies of two of the four selected ports unionize. The union from one of the ports is resistant to participating in the study for fear of performance problems that could hurt their bargaining power during future labor negotiations. However, corporate executives tell you they are willing to waive union responsibility for performance variances during the study, if their company is allowed to *choose* Proposal B, rather than being arbitrarily assigned to Proposal A or B by the researchers. The company has an incentive for assignment to Proposal B in order to take advantage of high-speed computers recently installed at the terminals. They believe that Proposal B will be selected over Proposal A by the Department of Homeland Security, and their terminals would then be in a superior position to compete for future TSA grants.

Construct a nonequivalent group design reflecting these changes to the study sample. Discuss threats to validity associated with this design, and ways to mitigate these problems (for example, use of cohorts, modifying the number and frequency of observations to minimize biases such as novelty effects and the effects of pre- and post-tests).

Exercise 2a

While working for the U. S. Government Accountability Office (GAO), you are assigned to conduct a program evaluation of the U. S. maritime drug interdiction program. By statute, the Coast Guard is the lead agency for maritime drug interdiction, supported by the Bureau of Immigration and Customs Enforcement (ICE). ICE is the lead agency for interdicting suspected drug flights over maritime transshipment areas. Generally, the effectiveness of the maritime drug interdiction program is judged by the amount of drugs intercepted relative to the estimated “flow rate” of drugs into the United States, as compared to the effort required (ship, aircraft, and boat patrol hours) for the interdictions. The Coast Guard and ICE operate a variety of ships, boats, aircraft, command and control installations, and intelligence networks to support the drug interdiction program.

Initial contacts with the Coast Guard and ICE headquarters staff in Washington, D.C., provide a bounty of information, which your assistant has compiled into two tables (Table 1 and Table 2).¹ However, variances in terminology, potential gaps and overlaps in data, and significant staff turnover among personnel at each agency concern you that relying on the data as presented, absent additional verification and “cleansing,” may lead to a faulty evaluation.

Provide a brief summary of potential problems in using the data from Table 1 and 2 to properly evaluate the effectiveness of the maritime drug interdiction program. Include issues of missing and incomplete data, data aggregation, unknown and changing data definitions, and data linked across time and clients.

Exercise 2b

Using the above scenario, you are in a situation where after identifying weaknesses in agency record data during the maritime drug interdiction program evaluation your team begins to implement a plan to overcome data sufficiency shortfalls. While interviewing ICE officials to clarify the large decrease in ICE tonnage intercepted between 2000–2001 and 2002–2003, you are advised that the definition for this category may have changed, reflecting an attempt by drug control policy makers to avoid possible double-counting by agencies involved in the same interdiction event. Similarly, you learned that the seemingly paradoxical decrease in patrol effort but heightened interdiction rate could be

¹ These are fictitious data, meant only for use in this exercise.

Supplement to the *Handbook of Practical Program Evaluation*, Second Edition, edited by Joseph S. Wholey, Harry P. Hatry, and Kathryn E. Newcomer, ISBN 0-7879-6713-0, cloth, © 2004 Jossey-Bass, An Imprint of Wiley.

explained by post September 11, 2001 mission changes and the advent of improved interdiction technologies.

Develop a cooperative plan with ICE and the Coast Guard to verify agency record data. Are there personnel at the respective headquarters offices that can bridge data gaps and overlaps? How about at the field unit level? Consider the need to review actual case files to expand on and verify the data you currently have, and the potential for access problems (such as classified information) and lack of agency personnel resources to assist your team. How might good planning mitigate these problems? How might you handle issues of missing or incomplete data and data inconsistencies and differing interpretation of results by each agency?

Table 1

U. S. COAST GUARD AGGREGATED DRUG INTERDICTION DATA

Year	Tons seized (all sources) destined for U.S.	Intercepts, but no seizure	Tonnage seized, by area	Estimated interdiction rate	USCG interdictions (tons)	Other Interdictions (tons)	USCG Patrol effort per ton (cost in \$ millions)
2000	175	45	PAC: 85 LANT: 90	15%	125	50	2.5
2001	210	35	PAC: 120 LANT: 90	15%	160	50	2.2
2002	200	25	PAC: 125 LANT: 75	25%	160	40	2.0
2003	200	20	PAC: 110 LANT: 90	25%	162	38	2.0

Note: Years in fiscal years.

Table 2

IMMIGRATION AND CUSTOMS ENFORCEMENT DRUG INTERDICTION

DATA

Year	Tons seized (all sources) destined for U.S.	Verified ICE pursuits, by area	Estimated flow rate (metric tons)	Tonnage seized, by area (metric tons)	ICE tonnage intercepted	Other tonnage intercepted	Interdiction rate per unit of effort (patrol hours)
2000	175	JIATF E: 60 JIATF W: 48	1050	JIATF E: 95 JIATF W: 85	85	90	0.05
2001	210	JIATF E: 60 JIATF W: 48	1100	JIATF E: 95 JIATF W: 110	88	122	0.08
2002	200	JIATF E: 60 JIATF W: 48	1100	JIATF E: 89 JIATF W: 110	22	178	0.08
2003	200	JIATF E: 60 JIATF W: 48	1100	JIATF E: 95 JIATF W: 115	24	176	0.08

Note: Years in calendar years.

Contributor: John Kennedy