

Chapter 1 Program Evaluation: An Introduction

Organizations, like human-beings and other creatures, must continually learn and adapt to survive and thrive. This process requires the creation of intellectual capital and its management to transform it into organizational intelligence. Thus, organizations which learn, thrive (i.e. achieve their mission and vision); those organizations which do not or cannot learn, die.

Organizational leaders and managers must make decisions; accordingly, they must gather, analyze, interpret, and apply information within a context influenced by values, laws, government policy, strategic and tactical objectives, workforce composition (changing demographics, knowledge bases, and skill sets), a hyper-competitive marketplace, and “speed of light” technological transformations. All of these activities must be accomplished quickly or an organization risks its agility and compromises its very survival. The authors argue that evaluation research will enable leaders and managers to efficiently and effectively engage these activities, while considering these multiple influences, leading to high quality decision-making. Rossi, Lipsey, and Freeman (2004, p. 370) write, "Evaluation involves more than simply using appropriate research procedures. It is a purposeful activity, undertaken to affect the development of policy, to shape the design and implementation of social [and other] interventions, and to improve the management of social [and other] programs...evaluation is a political [and technical] activity."

In this chapter, we examine an organizational learning model, the role of evaluation research in fostering intellectual capital, and program evaluation (three approaches).

I. The Learning Organization

A. On the Need to Learn

1. As organizations compete in the 21st century, the conventional wisdom in leadership and organizational development circles is that organizations which fail to continually learn so as to consistently improve efficiency, effectiveness, decision-making, product or service quality, and customer service are doomed to die (Deming, 1986, DiBella, 2001; Liebowitz, 2000; Juran, 1989; Senge, 1990).
2. Peter Senge (1990, p.3) in *The Fifth Discipline: The Art and Science of the Learning Organization* defined the learning organization as, “organizations where people continually expand their capacity to create the results they truly desire, where new and expansive patterns of thinking are nurtured, where collective aspiration is set free, and where people are continually learning how to learn together.”
3. According to Senge (1990, p.10), team learning is essential as teams are the “fundamental learning unit in modern organizations.”
 - a. If teams don’t learn, organizations don’t learn.
 - b. Teams are composed of individuals. If individuals don’t learn, neither do teams.
 - c. Learning or knowledge is therefore created on both the individual and team levels.

- If Senge's argument is correct that the future belongs to organizations which continuously learn and apply that learning to improving organizational culture and performance, then the process for creating organizational intelligence must be understood and mastered.

B. Organizational Learning

(Intellectual Capital + Knowledge Management = Organizational Intelligence)

- The process of generating organizational intelligence is presented in Figure 1.1. When intellectual capital is properly managed, an organization's intelligence is raised; thus, the organization learns.
- Intellectual Capital is (a) the purposeful acquisition of knowledge, (b) ability to apply that knowledge, and (c) the reservoir of experience in applying knowledge to (a) develop, research, and test solutions to organizational challenges; (b) capitalize on opportunities; and/or (c) enable the organization to engage in a healthy evolution towards improved efficiency and effectiveness. Sources of intellectual capital sources include schooling, the home, job or career related experience, self-directed study, team work, research, etc. Another source of intellectual capital is collateral (i.e., unintended) learning; collateral learning just happens.
- Knowledge Management is the combination of an organization's culture (political, economic, sociological, and psychological), administrative processes, information technology (IT) resources, and IT competency, which enables individuals, work teams, and organizations to create, sort, store, and retrieve intellectual capital for specific purposes or applications.

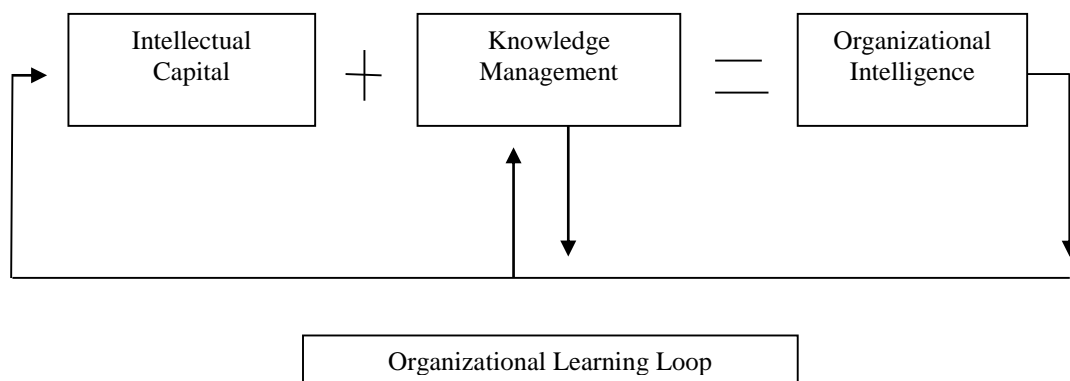


Figure 1.1 Organizational Learning Model

- Organizational Intelligence (i.e., its stock or cache of intellectual capital) is raised when intellectual capital is properly mediated or managed (knowledge management). Evidence of organizational intelligence includes an organization's patents, process and product innovations, performance (i.e., efficiency and effectiveness), stakeholder satisfaction, and cultural quality. Purposeful investment (e.g., training and technology) in fostering an organization's intellectual capital formation and its

knowledge management infrastructure will raise the organization's intelligence quotient.

5. Organizational Learning Loop contributes to the creation of intellectual capital and efficient and effective knowledge management when prior learning is purposefully linked to a continuous learning cycle. The "loop" takes the form of reports, conversations, presentations, policies, procedures, etc.
6. Accepting Figure 1.1 as accurate, leads one to conclude
 - a. That a high organizational intelligence "quotient" is the key competitive advantage for individuals (i.e., work teams) and organizations (e.g., businesses, schools, governments, non-profits, NGO's, etc.) to not only survive, but thrive in current business and professional environments.
 - b. Highly intelligent organizations will survive and thrive; unintelligent organizations will perish. Highly intelligent organizations continuously create and utilize intellectual capital.
 - c. The same fate awaits professionals (e.g., accountants, marketers, teachers, etc.) and managers (e.g., business leaders, assistant and full principals) who fail to remain current in their respective practices or disciplines and who are well advised to form learning alliances in order to continually produce intellectual capital.

C. The Role of Applied Research in Fostering Intellectual Capital Formation

1. Knowledge is created, discovered, or developed, either intentionally or unintentionally. The purpose of evaluation research is to intentionally create, discover, or develop knowledge. Evaluation research training teaches one to purposefully contribute to the formation of an organization's intellectual capital and to objectively critique research reports in the literature as well as improve decision-making.
2. Evaluation research incorporates a continuous quality improvement perspective which is embodied in the Shewhart or the PDCA Cycle (Deming, 1986, p. 88), as seen in Figure 1.2 which is a convenient summary of the evaluation research process.
 - a. In "P" (for plan) a problem is identified and solutions posited.
 - b. In "D" (for do) a small scale field-test is implemented; followed by "C" (for check) where the field-test experience and its effect is monitored and adjusted to achieve desired result.
 - c. Finally, in "A" (for act) the solution is implemented system-wide to achieve its full benefit, provided it produced the desired results; those affected are involved in the implementation.

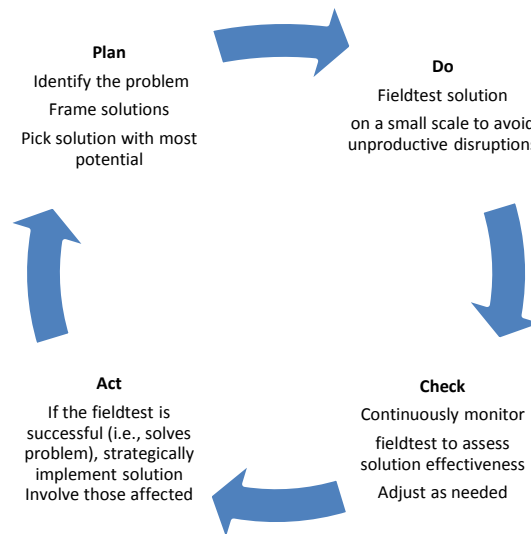


Figure 1.2 Shewhart Cycle

3. When we compare the scientific method to the Shewhart Cycle, we see that a similar process is employed. The scientific method typically includes these steps:
 - a. A problem, circumstance, or other question is identified and is deemed worthy of further investigation.
 - b. The problem, circumstance, proposition, or question is fully described (i.e., defined). This is done through a combination of expert knowledge, literature analysis, and professional judgment. Factors (often called variables) which either influence or are influenced within the context of the problem, etc. are identified.
 - c. Either research questions or hypotheses are formed next. This involves the drafting of variable operational definitions which guide what is to be measured by the researcher.
 - d. Next, a strategy for conducting the research is designed, guided by either by a research question, hypothesis or a combination of each. There may be multiple questions and/or hypotheses.
 - e. Data are gathered guided by the research question(s) and/or hypotheses.
 - f. Data are analyzed and reported.
 - g. The study is replicated to confirm earlier findings or/to assist other interested researchers.

4. Similar to improvement, evaluation requires the rendering of a judgment as to merit and worth. Merit refers to justified need, cost, or benefit. A judgment as to worth typically involves resource investment. Both judgments are influenced by values, technical knowledge, political and social considerations, and philosophy. For example, most reading improvement programs have merit, i.e., the purpose to improve reading achievement but a decision must be made as to which proffered program is worth the resource investment, given the desired outcome. Usually through a formal review process, a decision is made, which may be influenced by a single or combination of the factors mentioned above. Non-technical factors always

influence evaluative studies and resulting decisions; the study process, product (i.e., the report), and decision must be acceptable to stakeholders, especially those individuals, organizations, and communities affected.

5. Applied Research
 - a. Basic research is intended to “learn for learning’s sake,” develop new theory or refine existing theory. Often, the benefit or impact of basic search is not realized for years. Basic research is an investment, rarely with immediate payback.
 - b. Applied research assesses impact or seeks to resolve problems which affect real people or organizations.
 - (1) Applied research may be based on an established theory, emerging theory, or no theoretical framework at all. What is important is that the research is intended to have immediate or near-immediate “real world” application.
 - (2) Applied research has limited generalizability, i.e., findings are not able to be generalized beyond the population, sample, or target group studied.
 - c. Applied Research Studies: Types
 - (1) Descriptive studies are the most common form of action or managerial research. A circumstance or situation is described. Based on that description, decisions are made to effect change or not. For example, a cross-sectional study is conducted at a specific point in time to describe a specific condition or circumstance at a particular point in time. Longitudinal studies measure the dependent variable more than once, usually at pre-designated intervals over a specified time period.
 - (2) Exploratory studies are conducted so that decision-makers can more fully understand and/or comprehend a circumstance, problem, or situation. This enhanced understanding will hopefully lead to effective decision-making. For example, an exploratory study is conducted to identify issues, topics, etc. for further research and to provide data as to whether or not more formal research is warranted. Some authors suggest that exploratory studies are more lax in their requirements than formal research; however, all research, exploratory or formal, should be conducted rigorously.
 - (3) Predictive studies are those that forecast trends based on prior experience and key assumptions which are fed into statistical models.
 - (4) In an ex post facto research study, the independent variable is identified and then its presumed effect on the dependent variable is traced back through time; ex post facto research is conducted where it would be illegal, impossible, or unethical to manipulate the independent variable(s). In an experiment, the independent variable or variables and dependent variable are identified a priori and the independent variable’s effect on the dependent variable, observed as the experiment is conducted.
 - (5) An action research project is conducted “by people who want to do something to improve their situation, [e.g., a teacher who wants to improve classroom achievement]” (Sagor, 1992, p. 7). Results publication or study replication is not an objective. Sagor (1992, pp. 10-11) outlined the collaborative action research process (steps 1-5):

- (1) Frame the problem
- (2) Collect data
- (3) Analyze data
- (4) Report results
- (5) Produce an action plan to correct or reduce problem effect.

The process reads very much like the Shewhart or PDCA cycle. Arhar, Holly, and Kasten (2001) offer an excellent discussion of action research.

6. Evaluation Research as Applied Research

- a. Evaluation research is aimed towards providing information to make efficacious decisions to improve programs and services, establish accountability, and/or to generate new knowledge. Evaluation research application is immediate and "real world" usually with little or no generalizability beyond the customer/client target group studied.

b. Two Primary Categories of Applied Evaluation Research

- (1) Formative evaluation studies are conducted during the "life-span" of an intervention program (e.g., a reading improvement or continuous quality improvement effort) or other term-limited purpose (e.g. a marketing or fund raising campaign) to assess effectiveness. Often, mid-course corrections are made to ensure that the desired outcome is attained (Scriven, 1991). Within the evaluation research context, there are typically three types of formative evaluation studies.

- (a) Needs Assessment studies determine whether or not a particular service, cluster of services, product, etc. is needed by customers/clients in order to accomplish the intended goals and objectives of an organization so that it will achieve its mission and vision, and abide by its values.

- (b) Implementation Evaluation studies may be cross-sectional or longitudinal and are conducted in order to determine whether or not a program, [e.g., service, cluster of services, product, etc.] is being implemented as designed.

[1] Cross-sectional implementation evaluations are conducted, usually once, at a specific point in time, e.g., during the 12th month of a 36 month service initiative to foster midcourse adjustments to enable the program to more efficiently and effectively achieve its goals and objectives.

[2] Longitudinal implementation evaluations are conducted more frequently as a form of continuous implementation evaluation, where at specific points in time, measurements are taken regarding "current" program implementation and impact. This program or progress monitoring enables "continuous" program adjustment so that goals and objectives are more efficiently and effectively achieved.

- (2) Summative (or impact) evaluation studies are conducted after the completion of an intervention program or other term-limited event. Such studies answer the question, "What was the impact of the event?" or "Was the desired outcome attained?" (Scriven, 1991).

- c. Formative and summative evaluation research employs the same research processes, study designs, data collection strategies, data analysis, and reporting conventions as does any other type of applied research.

II. Program Theory Driven Evaluation

A. Introduction

1. Stufflebeam (2001) has identified 22 different approaches to evaluation; the focus in this primer is on program evaluation. Rossi, Lipsey, and Freeman (2004, p. 32) argue that a "good evaluation design is one that fits the circumstances while yielding useful answers to the questions that motivate it."
2. Evaluation studies must be tailored to each unique evaluative context. A program evaluation question or questions (Chapter 2) must be posed and agreed upon between the evaluator, the evaluation team, and relevant stakeholders. Using the framework provided by Rossi, Lipsey, and Freeman (2004, p. 54), evaluation questions fall into five categories. Table 1.1 is a comparison of various evaluation classification study schema.
 - a. In a needs analysis, the evaluation question is diagnostic in that it seeks to determine whether or not there is a need for a program. Needs analyses describe a problem or opportunity, those affected, the consequences to those affected, if the problem persists or the opportunity is unrealized, and the benefits to those affected if the problem is ameliorated or opportunity realized.
 - b. In a program theory assessment, the evaluation question focuses on how the program is conceptualized and designed. Comparison is made to similar programs in the form of the literature review or synthesis which is drawn from the relevant professional and empirical literature.
 - c. In a program process assessment, the evaluation question focuses on the quality of current program operations, examining program management, program efficiency, service efficacy, etc. Program process assessment is typically an element of implementation evaluation.
 - d. In an impact assessment, the evaluation question examines what effect or impact a program has had at specific points in time during the "life" of the program (implementation evaluation) or after the program has been terminated (summative evaluation).
 - e. An efficiency assessment is similar to a study of return on investment or cost-benefit analysis. Such studies require specialized financial expertise as program outcomes are compared to their costs to determine merit and worth.
3. Planning considerations include formulating the evaluation question(s); selecting appropriate evaluation methods; mapping, fostering, and preserving stakeholder and team relationships; and understanding the program space, management, theory, and resources.
 - a. First, the program to be evaluated must be understood by the evaluator and the evaluation team. This requires a thorough examination and understanding of stakeholder relationships and the program's theory, goals (distal outcomes), objectives (proximal outcomes), customers/clients, and methodology.

Table 1.1
Program Evaluation Study Classifications

Broad Category	Specific Category	Alternative Category	Purpose
Formative	Needs Analysis	Needs Assessment	Establish program need
Formative	Program Theory Assessment	Program Theory	Establish program theory veracity
Formative	Program Process Assessment	Cross-Sectional or Longitudinal Implementation Evaluation	Document actual program implementation
Summative	Impact Assessment	Impact Evaluation	Document final program impact
Summative	Efficiency Assessment	Cost-Benefit Evaluation	Document program Return on Investment (ROI)

- b. Second, The evaluator must understand (map) the complex web of relationships between stakeholders; critical stakeholders must be continuously informed about the program evaluation. Conscious, authentic effort must be invested in fostering and preserving relevant important relationships between and among stakeholders. Such healthy relationships can prevent and resolve issues which may impede the progress of the program evaluation. Further, healthy relationships between stakeholders and especially decision-makers, will increase the likelihood that evaluation results will be utilized in decision-making.
- c. Third, Any evaluation plan should specify who gets what information and when, including any interim and final reports by the evaluator.
- d. Once steps a-c are completed, the evaluation study must be designed.
 - (1) Design decisions are influenced by the study's purpose as discussed above; the program's developmental stage (e.g. proposed programs are likely to undergo a program theory assessment; new programs are likely to receive a form of implementation evaluation; and mature programs are likely to receive an impact evaluation or effectiveness assessment so that decisions on program continuance or termination can be made); program evaluation management expertise (the evaluator and evaluation team must have the technical and social competence to successfully conduct and complete the evaluation); and program evaluation resources (there must be sufficient human, material, and financial resources to successfully execute the study).
 - (2) For guidance on evaluation study design, see Chapters 7, and 8.
- e. Once steps a-d are achieved, data collection strategies (measurement devices) are determined; these will typically involve either the construction or purchase of measuring instruments. Purchase or construct after the contract is signed.
 - (1) All measuring devices must at a minimum be reliable and valid (Chapter 3).

- (2) If a survey instrument (e.g., questionnaire), rating or ranking scales, or an attitudinal index are to be constructed, see Chapter 4.
 - (3) If an achievement test (measures learning) or a performance assessment (measures skill demonstration) is to be constructed, see Chapter 5.
 - f. Once steps a-e have been accomplished, a data analysis plan is prepared; for guidance, see Chapters 9-13.
 - g. Once steps a-f have been completed, usually an evaluation research proposal is prepared; this becomes the basis for a negotiated evaluation contract or agreement. For guidance on constructing an evaluation research proposal, see Chapter 6, The Evaluation Research Proposal.
 - h. Once steps a-g have been completed and the evaluation research proposal or agreement has been successfully negotiated, then evaluation study is conducted.
 - i. Once steps a-h have been completed, an evaluation report must be prepared; for guidance see Chapter 6, The Evaluation Research Report.
4. Important Terms/Concepts
- a. A program may be defined as a single service, cluster of services, product line, an organizational strategy, etc. The definition of “program” is very flexible. It is the responsibility of the evaluator or evaluation team to clearly describe the "program" under examination so that a reasonably informed reader will understand what is being examined. A program evaluation may be simple (e.g., determining which box of cereal to buy given a balance between cost, nutrition, and taste preference) or complex (e.g., determining the most effective strategy [i.e., whole language, literature based, or balanced literacy] for teaching reading). Program evaluation studies should produce timely, accurate information which is understood and useful to end-users who are typically stakeholders and decision-makers.
 - b. Rossi, Lipsey, and Freeman (2004, p. 16) define program evaluation as, "the use of social research methods to systematically investigate the effectiveness of social [or other] intervention programs in ways that are adapted to their political and organizational environments and are designed to inform social [and other] action to improve social [and other] conditions." An evaluation sponsor is the individual or the organization which commissions the evaluation study; this person or organization may or may not be an evaluation stakeholder.
 - c. A stakeholder is any individual or organization which has a significant interest in the evaluation study outcome, program operations, benefits, etc.; these may include organizational leaders, program managers, program staff, suppliers, customers/clients or anyone else who has a “stake” (interest) in the organization, program, or evaluation study. Stakeholders usually present with competing interests, varied perspectives, and differing reactions; all of which must somehow be accommodated through negotiation in order for the program evaluation to be successfully conducted. The program evaluator should remember that stakeholders are primarily loyal to their constituencies; thus, conflicts are inevitable and should be resolved through negotiation, using available resolution resources and procedures. Some advocate that the program evaluator be involved

in negotiating between and among stakeholders. Others advise against this. Perhaps the most cogent advice is that the individual, ideally in the appropriate role, who is best equipped to mediate and negotiate conflict resolution, is the one who does so. The ideal mediator is likely to change as stakeholders and issues change.

- d. An evaluation team or committee designs and conducts the evaluation study, including the reporting and dissemination of results. The evaluation team comprises the lead evaluator (or principal investigator), co-evaluators (co-principal investigators), the evaluation sponsor, and the most relevant stakeholders who are capable of contributing to the evaluation effort.
- e. Program theory is an explanation of how the program is/was intended to achieve its goals and objectives. This description is graphic with narrative explanation. A distal outcome is a critical or key goal(s) of the program. Measuring distal outcomes may not be possible during the “life span” of the program as they are often expressed in lofty unmeasurable terms. If it is not possible to measure a distal outcome, then proximal outcomes substitute. Distal outcomes may also serve as a key program performance standard against which actual program performance is compared to determine actual performance sufficiency and acceptability. A proximal outcome is a program objective. The logical relationship between a proximal outcome and a distal outcome is that proximal outcomes must be achieved in order to achieve a distal outcome. An action step is a sequential series of actions necessary to achieve a proximal outcome.

B. Important Evaluation Principles

Program evaluation is both a political, technical, and financial event. Within organizations, as within society, there are typically honest disagreements over program purpose, goals, objectives, allocated resources, benefits, and opportunity costs. Evaluation also portends potential change which is often feared and passively or actively resisted. Evaluation tensions may center on several issues:

1. Trade-offs: Program Evaluation Design and Stakeholder Resistance
 - a. The evaluator must recognize that there may be tension between the technical demands of the program evaluation and the “reality” within which the evaluation is to be conducted. Resolution of these tensions must be negotiated by the evaluator, stakeholders, and/or the evaluation sponsor.
 - b. Evaluation acceptance may be enhanced by including appropriate stakeholders on the evaluation team or evaluation study supervision committee. Frequently, the key to reducing resistance is most often meaningful participation in the process. Stakeholder program evaluation participation and the parameters of that participation must be negotiated between the evaluator and the evaluation sponsor.
 - c. Sometimes, it is necessary to trade some evaluation design rigor or adjust (add, modify, or delete) evaluation research questions to “buy” stakeholder participation. Ensure to the practicable extent possible that “the things that matter most” are not traded away.

- d. Any trade-offs must be approved by the evaluation sponsor, who is usually paying for the evaluation, and meet ethical evaluation standards.
2. Program Evaluation Volatility
 - a. Organizational (as do societal or government) priorities, policies, and interests change. These changes may result in program evaluation study elimination, an unexpected urgency in communicating either preliminary or final evaluation findings, or changes in the real or perceived program evaluation stakeholder risk profile. The more volatile the program evaluation context is; the greater may be the degree of difficulty in successfully completing a suitably rigorous evaluation study.
 - b. The evaluator or evaluation team must be in honest, open communication with the evaluation sponsor and stakeholders, seeking to negotiate a suitable resolution or accommodation to any presenting issue or problem. Usually, the evaluator has little influence or control over the program evaluation context; he or she must above all else be ethical in his or her conduct.
 - c. The evaluator and/or evaluation team must anticipate potential volatility and develop contingency plans to accommodate or adapt to volatile program evaluation contexts as the purpose of any evaluation study is to answer the evaluation questions so that efficacious decisions can be made.
 3. Program Evaluation Ecology
 - a. Every program evaluation study regardless of simplicity or complexity has effects (which may be perceived as consequences), affects stakeholders (positively or negatively depending on perspective), and operates in unique contexts (political, economic, cultural, demographic, and linguistic).
 - b. In order for an evaluation study to be conducted, program evaluation ecology must be identified, understood, and successfully navigated. This requires the evaluator to be extremely familiar with the program being evaluated; it's stakeholders; it's political, economic, cultural, and demographic contexts; and suitable evaluation designs, data analysis tools, and analytical frameworks in order to produce accurate information for efficacious decision-making.
 - c. Communication and demonstrating understanding of the various dimensions which comprise a program evaluation's ecology is critical to successfully generating stakeholder buy-in and support, conducting and completing the evaluation study, ensuring the acceptability of evaluation findings, and getting the results used for decision-making.
 - d. Strategies for successfully navigating a program evaluation's ecology include early and appropriate stakeholder involvement; frequent, open communication which is understood by stakeholders; and adherence to ethical standards.
 4. Program Evaluation Results Utilization
 - a. The hope of any evaluator is to have the results of his or her study considered within decision-making. While there are a host of factors that go into decision-making, a high quality politically acceptable evaluation study will materially contribute to efficacious decision-making.

- b. There are several factors under the control of the evaluator or team that will positively influence utilization prospects. Typically, there are two reports the evaluator or committee writes. The technical report goes primarily to the evaluation sponsor and is similar to the executive summary, but delves into much more detail. The technical report is known as primary dissemination. The executive summary of the report is written primarily for stakeholders and decision-makers. The executive summary is an abbreviated report of the study's commission, methodology, the program goals (distal outcomes) and objectives (proximal outcomes), degree of goal and objective achievement, important findings, conclusions (including a rationale), and recommendations. The executive summary is known as secondary dissemination.
 - c. The effective evaluation report has a few critical characteristics. The evaluation report should be written in jargon free language so that it is easily understood by a reasonably informed stakeholder or decision maker. The evaluation report should be reviewed and approved by the evaluation sponsor as well as representatives of key stakeholder groups. The report should be comprised of a judicious and purposeful mix of figures (charts and graphs) and narrative. The report should be easy to read, graphically appealing, and as brief as is possible. Detailed information should be presented in an appendix as opposed to the main body of the report.
5. Program Evaluation Ethics
- a. The evaluator or team should involve, to the extent permissible and ethical, decision-makers in the design, execution, and completion of the evaluation study. Ideally, this will build trust between the evaluator, stakeholders, and decision-makers in the integrity of the evaluation process and the information it produced. Program evaluations must always conform to the highest applicable professional ethical and legal standards. McMillan and Schumacher (2006, pp. 142-145) outline several principles for conducting ethical research which are summarized below.
 - (1) The lead researcher (also called primary investigator) must ensure the ethical treatment of all study subjects, human or animal, and compliance with all controlling national government, federal/provincial, local government, and organizational regulations and policies.
 - (2) Study subjects and any parents or guardians (if subjects are minors) should be honestly informed as to the study's purpose, benefits, risks (physical or psychological harm or discomfort) and time frames.
 - (3) As a general rule study subjects (if of legal age) or parents or guardians should sign informed consent agreements. Frankel and Wallen (2006, p. 57) suggest that informed consent agreements encompass an explicit consent to participate, acknowledgment of being informed of any risk or risk potential, a statement by the subject of the risks or risk potential in his or her own hand, participation conditions (e.g., voluntary participation and right to "quit" at any time) and a general hold harmless clause.
 - (4) Subject information (e.g., information that could potentially identify a subject) must be kept confidential, unless disclosure is required by law, e.g., knowledge of child abuse that is acquired by a researcher.

- b. Most organizations (i.e., school districts, colleges, universities) and many non-profit and for-profit organizations maintain institutional review boards (IRB). The IRB must review and approve research projects, including dissertations, conducted on its property or under its auspices. McMillan and Schumacher (2006, p. 145) and Frankel and Wallen (2006, p. 56) report that most educational research involves little or no risk to subjects; it is therefore usually exempt from formal review processes, e.g., IRB. However, most organizations require IRB review and approval before university students, faculty, or their own employees begin a study process that is likely to produce a “formal” dissemination of results.
- c. The federal regulations governing human subjects’ research are found in Parts 45 & 46 of The Code of Federal Regulations (CFR). The American Evaluation Association has published Guiding Principles for Evaluators (retrievable from <http://www.eval.org>). These guiding principles consider five areas:
 - (1) Systematic Inquiry
 - (2) Competence
 - (3) Integrity/Honesty
 - (4) Respect for People
 - (5) Responsibilities for the General in Public Welfare
- d. The American Educational Research Association has promulgated ethical standards which may be retrieved at < <http://www.aera.net/AboutAERA/AERARulesPolicies/ProfessionalEthics/tabid/10200/Default.aspx>>. The American Psychological Association’s Code of Ethics may be retrieved at < <http://www.apa.org/ethics/>>. Most professional and trade organizations have published codes of ethics and research standards. Researchers are responsible for honoring their disciplinary codes as well as applicable national, state or provincial, local, and organizational codes of ethics, research standards, or law. Yarbrough, Shulha, Hopson, & Caruthers (2011) authored the third edition of the Joint Committee on standards for educational evaluation. Ethical evaluation studies conform to these standards. These standards are:
 - 1. Program Evaluation Management
 - 2. Human Subjects Protection
 - 3. Evaluation Methodology

C. Program Theory Evaluation: Explanation

- 1. Every program, regardless of how it is defined, has a rationale or plan for its operation, i.e., how it is supposed to work so that its goals (distal outcomes) and objectives (proximal outcomes) are accomplished. This rationale or plan is referred to as the program's theory.
 - a. Rossi, Lipsey, and Freeman (2004, p. 44) define program theory as, “[I]t’s plan of operation, the logic that connects its activities to the intended outcomes, and the rationale for why it does what it does.” A cogent, precise program evaluation theory makes it easier to not only understand how the program is intended to work, but also attribute intended and unintended effects (impacts) to the program, as the case

may be.

- b. The program theory must be defined and described by the evaluator (using program documents, staff and customers/clients interviews, and the professional and empirical literature) prior to conducting an evaluation.
 - (1) The description includes the program's underlying research base, program structure, management model (management structure and processes), services portfolio, service delivery systems, and intended outcomes.
 - (2) The program's internal structure and services are graphically linked to intended outcomes with supportive and explanatory narrative.
 - (3) There must be a self-evident logical relationship between all the parts of the program theory; this requires a thorough understanding of the evaluative context as well as relevant research and prevailing best practices on the part of the evaluator and several members of the evaluation team to describe and critique a program's theory.
- c. For simple programs, logic models tend to be easily constructed and understood. For more complicated programs, each major element of the program should have its own logic model constructed. It is better to work with a few simply constructed, but understood, logic models rather than "trying to fit everything" into one "monster" graphic. There are three (3) logic model graphics for each program theory explanation:
 - (1) The Service Utilization Plan
 - (2) The Program Organizational Plan
 - (3) The Program Impact Theory
 These graphic logic models are profiled in Figures 1.3 to 1.6.

2. The Service Utilization Plan

- a. The Service Utilization Plan Describes
 - (1) The rationale for selecting the specific services to be offered;
 - (2) The marketing strategy to inform program targets;
 - (3) The who, what, when, and where services will be provided; and
 - (4) How quality standards are maintained and advanced.
- b. The purpose is to initiate the change process to produce the intended benefit.
- c. The plan is written from the perspective of the programs intended beneficiaries.
- d. The service utilization plan can be described in a flowchart, supported by narrative. A simple flowchart for an afterschool reading program is depicted in Figures 1.3.

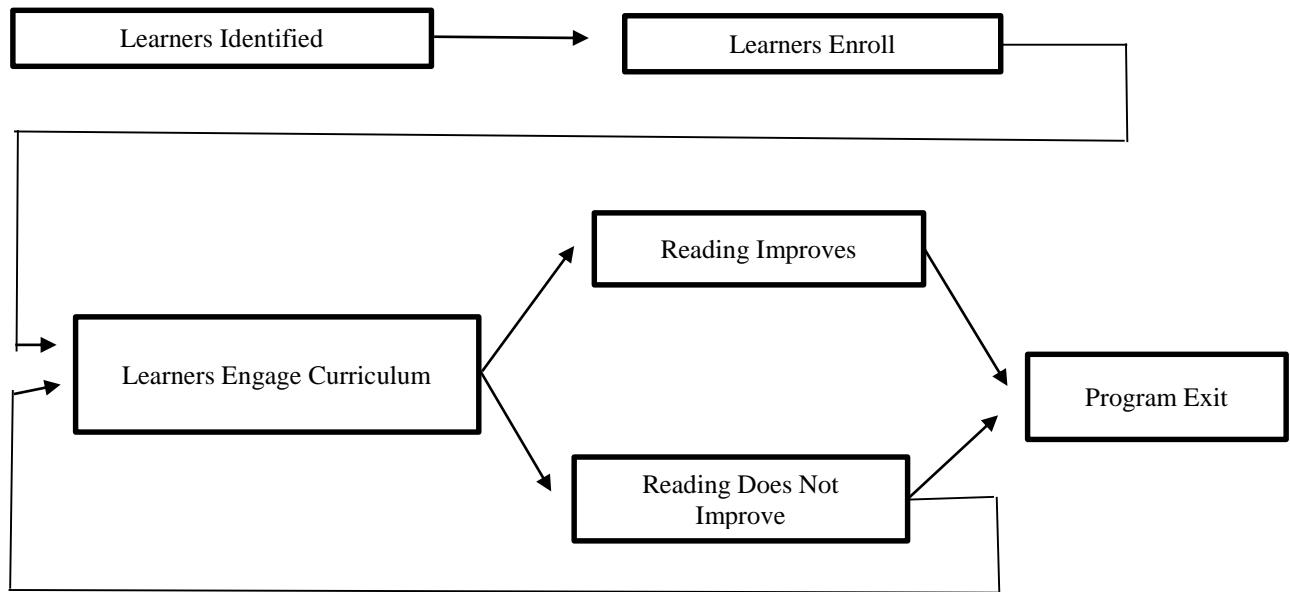


Figure 1.3 Program Service Utilization

5. The Program Organizational Plan (Same After-School Reading Program, Figure 1.4)
 - a. The school administration provides management support, funding, community/school relations, administrative evaluation, etc.
 - b. The school reading team designs the curriculum, identifies eligible students, obtains parental consent, trains teachers, selects curriculum materials, provide student rewards, evaluates the program, etc.
 - c. Within program teaching, the developmental reading curriculum is taught; students assessed, re-taught, and exit the program after achieving proficiency or for other reasons.
 - d. Transportation manages the process and means by which students are returned to their respective homes.

6. Program Impact Theory (Same After-School Reading Program, Figure 1.5)
 - a. This is the cause and effect element of the Rossi, Lipsey, and Freeman (2004, pp. 139-146) program theory framework.
 - b. The program impact theory identifies the causes (i.e., Service Utilization and Program Organizational Plans) which are expected to produce the desired impact on the target population.
 - c. Rossi, Lipsey, and Freeman (2004, pp. 140, 142) identified two types of program outcomes (or impacts).
 - (1) Proximal outcomes are more immediate, direct outcomes and are easier to attribute to the program.
 - (2) Distal outcomes are dependent on proximal outcomes, are more long-term, and more difficult to document and attribute to the program.
 - d. The Program Impact Theory is described in Figure 1.5.
 - (1) Program teaching is where the program/target transactions (teaching/learning

- reading skills) occur.
- (2) The proximal outcome in the model is higher reading scores which reflect improved reading skills and ability.
 - (3) The distal outcome is improved student achievement; reading is fundamental to academic success.

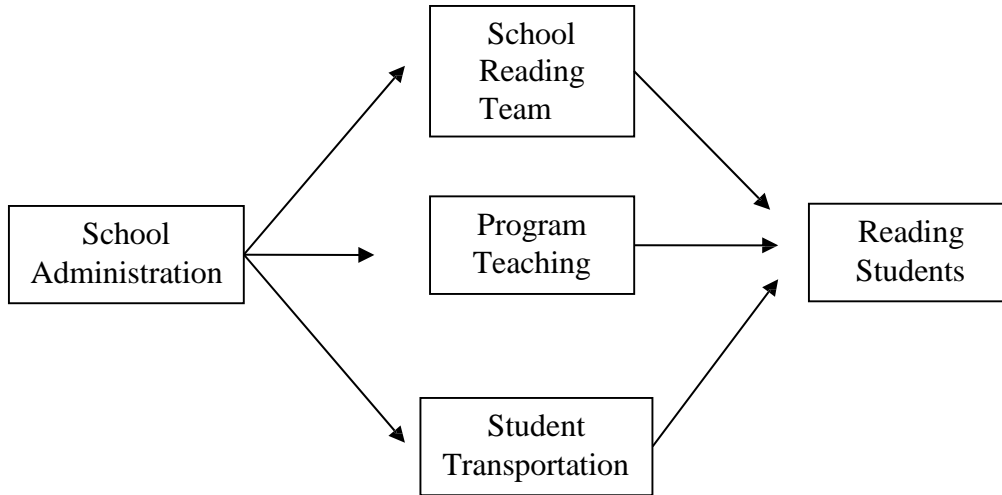


Figure 1.4: The Program Organizational Plan

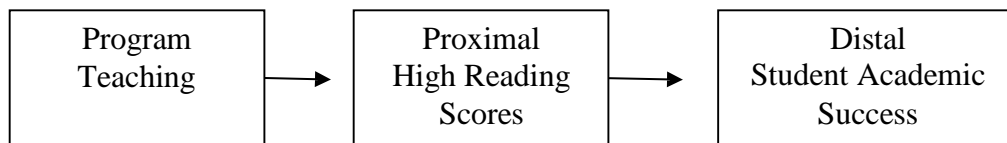


Figure 1.5 Program Impact Theory

7. The Program Process Theory (Figure 1.6) comprises:
 - a. Service Utilization Plan expresses the program's core assumptions, strategies for reaching intended clients, and describes how services are delivered or discontinued.
 - b. Program Organization Plan describes the program's administrative structure, staffing model, policies and procedures, funding, and quality control strategies.
 - c. Program Impact Theory is the combined effect on the program customers/clients, achieved through the combined service utilization plan and program organizational plan.

When merged, the program service utilization plan, organizational plan, and impact theory is referred to as its process theory (Rossi, Lipsey, & Freeman, 2004, pp. 139-46).

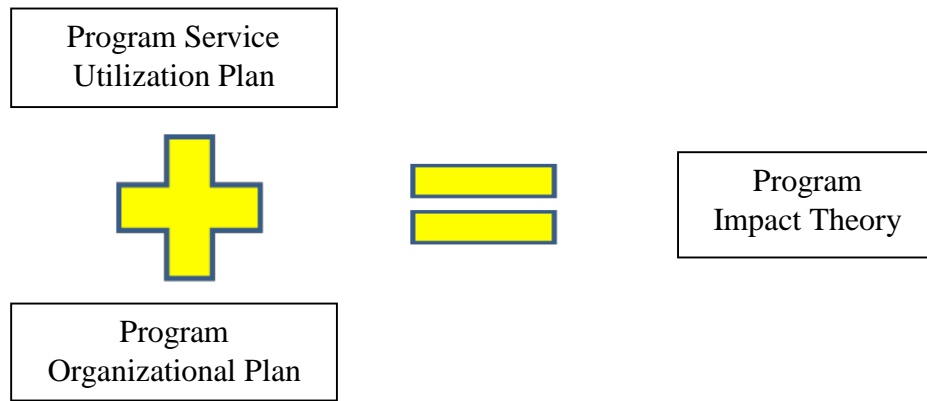


Figure 1.6 Program Process Theory

D. Strategies of “Uncovering” A Program’s Theory

1. Articulating a Program’s Theory
 - a. An evaluator must explicitly describe the program’s theory which may take one of these two forms:
 - (1) An articulated program theory is one that is clearly expressed and agreed upon by critical stakeholders.
 - (2) An implicit program theory is one that has not been fully and clearly described, requiring the evaluator to "dig for it."
 - b. When describing a program’s theory, the evaluator's purpose is to "described as designed," i.e., describe how the program was intended to produce its desired outcomes.
2. Framing Program Boundaries
 - a. Ask program decision-makers, who will act on the evaluation findings, to identify program components to be assessed (ensure that all critical elements, i.e., services, material and human resources, program processes, etc. are included.)
 - b. Work backwards from intended program impacts to identify all critical program components and processes, vetting your understanding of the program decision-makers and key stakeholders.
3. Describing the Program Theory
 - a. Once the boundaries of the program have been framed, Rossi, Lipsey, and Freeman (2004, pp. 148-152) advise the evaluator to describe the program.
 - (1) Working with program planners, managers, and stakeholders, the evaluator draws out (elicits) from them how the program was intended to achieve its anticipated impact.
 - (a) If the program is relatively new, the evaluator may be able to actually help a program planner frame the program theory.
 - (b) If the program is more fully developed or implemented, then the evaluator must describe the program theory based on information from program stakeholders, staff, clients, documents, and operations; however, the purpose is still to describe the program as it was intended to function.

- (2) The evaluator drafts a description and submits it to relevant stakeholders and the evaluation sponsor who reviews and comments; the process repeats until there is general agreement.
- b. Program Theory Information Sources
 - (1) According to Rossi, Lipsey, and Freeman (2004, pp. 149) there are four primary information sources to describe the program theory:
 - (a) Program documents;
 - (b) Stakeholders & Key informant interviews;
 - (c) Site visits to observe program operations; and
 - (d) The relevant professional and empirical Literature.
 - (2) From these four primary information sources, Rossi, Lipsey, and Freeman (2004, pp. 149-151) identify three types of information which are useful in describing program theory:
 - (a) Program goals and objectives,
 - (b) Program processes and services, and
 - (c) Logic linking intended outcomes with program processes and services.
 - c. Describing Program Goals and Objectives
 - (1) By examining program documents and/or conducting interviews, the evaluator can document intended program goals and objectives.
 - (2) Goals and objectives speaking to intended outcomes relate to the program's impact.
 - (3) Goals and objectives addressing program services speak to the program service utilization plan.
 - (4) Goals and objectives relating program management, funding, or staffing relate to its organizational plan.
 - d. Describing Program Processes and Services
 - (1) Each program process (mechanism for program management or producing services) must be identified and described.
 - (2) Each program service must be identified and related to its enabling program process and intended outcome.
 - (3) This information may be obtained from program documents, (e.g., job descriptions, interviews, and document reviews).
 - e. Logic Testing the Program Theory
 - (1) After the describing the program's theory, the evaluator must examine the logic and/or conceptual linkage between the program's organizational plan and services plan and intended outcomes, using judgment based on experience and research.
 - (2) These linkages are usually reduced to flowcharts, with analytical narrative which objectively examines the relationships graph for logic and conceptual accuracy.

- f. Seeking Consensus on the Program Theory Description
 - (1) Once the initial draft of the program theory description is completed by the evaluator, it should be submitted to a committee of knowledgeable persons for examination and corroboration.
 - (2) This is likely an iterative process, requiring multiple reviews until general consensus is reached that the program theory is accurately described.
 - (3) The corroborating committee should be composed of relevant stakeholders, program staff, customers/clients, and external members (who receive no program benefits).
 - (4) The outcome of this process should be a concise, accurate, and agreed upon description of how the program was intended to function and deliver anticipated impact on program targets (customers/clients).

E. Strategies for Assessing a Program's Theory

1. Once the program's theory has been accurately described, the evaluator must assess for feasibility implausibility (i.e., soundness).
 - a. When a program is based on sound theory, the results associated with the program may be more accurately attributed back to it. Thus, we learn not only the "how" of success but the "why" as well. An evaluator can also better explain program failure.
 - b. An assessment of the program's theory for feasibility and plausibility (soundness) requires the evaluator to examine what is known about the program from the needs assessment; the evaluator's stakeholders', and program staff's expertise and knowledge of the program; and the relevant professional and empirical literature, including the program's own documents.
2. Mechanics of Assessing a Program's Theory
 - a. The program service utilization plan and the program organizational plan should be examined separately.
 - b. The examination committee should include knowledgeable stakeholders, evaluators, decision-makers, external experts, the evaluation sponsor, etc.
 - c. The examining committee agrees *a priori* on the scope (i.e., what is examined and by whom) and sequence (i.e., what is examined in what order) of the program theory assessment.
 - d. "Look For's" when Assessing Program Theory
 - (1) Assess the theory for internally consistent logic and general plausibility considering its needs assessment and proximal and distal intended outcomes. Specifically examine:
 - (a) Assumptions about what services are offered and which are accessible to the target population;
 - (b) Service access barriers and strategies to overcome;
 - (c) The cultural "fit" between program services, staff, and the organization and customers/clients; and
 - (d) The managerial competence of the program to deliver the intended service plan to program customers/clients.

- (2) Compare the program theory description to the relevant professional and empirical literature, including prevailing best and emerging professional practices to:
 - (a) Compare to highly similar programs and/or
 - (b) Compare only those elements of the program theory for which relevant literature is available, if there is scant literature available.
 - e. Assess the program theory by observing it in action. Observe service delivery and interview participants (program staff and customers/clients).
3. There is an expected link between the program's theory and anticipated outcomes (e.g., proximal and distal outcomes). The examining committee assesses whether or not this causal link or chain is plausible given the needs assessment, available program performance evidence, and intended outcomes.

III. Tyler's Goals Oriented Evaluation Process and the Kirkpatrick Evaluation Framework

A. Tyler's Goal Oriented Evaluation Process

1. Tyler's (1949) Goals Oriented Evaluation Process
 - a. Learning objectives or targets are identified and expressed in behavioral terms.
 - b. Benchmarks or attainment indicators for each learning objective or target are expressed in behavioral terms.
 - c. Relevant content is identified and organized to align with learning outcomes and benchmarks or learning targets and attainment indicators.
 - d. Instructional activities are designed to enable students to learn the specified knowledge and skills, and/or exhibit desirable attitudes.
 - e. Instruments (e.g., achievement test or direct performance assessments) are developed to measure student attainment (mastery) of the learning outcomes or targets.
 - f. Once measuring instruments are administered and data analyzed, instructional adjustments for current students (e.g., re-teaching or moving on to the next topic, if any) are made or the learning outcomes and curriculum may be revised.
 - g. The process repeats in a continuous cycle for continuous improvement.
2. Tyler's (1949) evaluation framework is widely used in education and training. It is the methodological foundation for Chapters 4 and 5 in this primer.

B. Kirkpatrick's Evaluation Framework

1. Kirkpatrick's Evaluation Framework consists of four levels, which when combined presents a chain of evidence to argue for a program's effectiveness or failure (Kirkpatrick & Kirkpatrick, 2006; 2007). The Kirkpatrick evaluation framework is a powerful device for organizing the interpretation and synthesis of evaluation research results. The four levels are:
 - a. Reaction measures participant feelings, satisfaction, and opinions about the training event. Reaction is measured by brief, simple surveys often referred to as "smile sheets."

- b. Learning measures participant knowledge, skills, and/or attitudes as well as intended behavioral change as a result of participating in an educational or training event. Learning is measured by an achievement test or direct performance assessment. Intended behavior change is measured by self-report using a brief survey instrument as are attitudes.
 - c. Behavior measures participant behavior change after the training event “on- the-job” usually at specified points in time. Measuring actual behavior change involves the trainer or instructional designer, the participant’s supervisor and sometimes colleagues.
 - d. Results is the highest level in the Kirkpatrick evaluation framework. “The final results can include increased production, improved quality, decreased costs, reduced frequency and/or severity of accidents increased sales, reduced turnover, and higher profits” (Kirkpatrick & Kirkpatrick, 2006, p. 25). These results tend to be at the unit or organizational level.
2. The Kirkpatrick evaluation framework is the most widely used evaluation perspective in the world today.
- a. It consists of four levels ranging from Reaction to Results where each successive level builds upon the one before. Thus, in order to assess Level 4 Results, Levels 1, 2, and 3, must have been measured and documented. This approach is integral to building a chain of evidence for arguing education and training event effectiveness.
 - b. Measuring the Kirkpatrick Evaluation Framework Levels
 - (1) The item formats found in Chapter 4 can be used to construct measures of Level 1 Reaction.
 - (2) Achievement tests and direct performance assessments (Chapter 5) can be used to measure Level 2 Learning.
 - (3) Item formats found in Chapter 4 and direct performance assessments found in Chapter 5 can be used to measure Level 3 Behavior.
 - (4) Financial ratios such as those found in Chapter 9 can be used to measure Level 4 Results.
 - c. Kirkpatrick and Kirkpatrick (2006; 2007) provide extensive guidance on measuring each of the four levels; both of these books are worth the read for serious evaluators who practice primarily in education and training. Kirkpatrick Partners’ website provides substantial additional information <http://www.kirkpatrickpartners.com/Home/tabid/38/Default.aspx>

Review Questions

Directions. Read each item carefully; either fill-in-the-blank or circle letter associated with the term that best answers the item.

1. The phrase, “learning for learning’s sake,” refers to _____ research.
 - a. Basic
 - b. Applied
 - c. Replication
 - d. Descriptive

2. What is the fundamental learning unit of the modern organization?
 - a. Individuals
 - b. Teams
 - c. Departments
 - d. Divisions
3. What creates intellectual capital?
 - a. Organizational Intelligence
 - b. Knowledge management
 - c. Organizational learning
 - d. HR investment
4. A research problem, issue, opportunity, etc. is significant if its results
 - a. Provide knowledge about best or enduring professional practices
 - b. Develop a new or revise an existing theory
 - c. Extend knowledge
 - d. All of the above
5. A research study is generally based on potential contribution to
 - a. General knowledge
 - b. Theory
 - c. Professional practice
 - d. All of the above
6. The definition, “collecting data from a sample at different points in time in order to study changes or continuity in the sample’s characteristics” is most likely to refer to which type of study?
 - a. Descriptive
 - b. Longitudinal
 - c. Causal-comparative
 - d. Cross-Sectional
7. This study where a researcher can examine cause-and-effect relationship(s) of a(n) independent variable(s) on a dependent variable where it would be illegal, impossible, or unethical to manipulate the independent variable(s) is:
 - a. Descriptive
 - b. Ex Post Facto
 - c. Longitudinal
 - d. Cross-Sectional
8. The study which measures the characteristics of a sample at one point in time is:
 - a. Descriptive
 - b. Ex Post Facto
 - c. Correlation
 - d. Cross-Sectional
9. If your research study requires you to search for relationships between and among variables, your design is said to be:
 - a. Descriptive
 - b. Longitudinal
 - c. Correlation
 - d. Cross-Sectional

Selected Answers: 1. a, 2. b, 3. c, 4. d, 5. d, 6. b, 7. b, 8. d, & 9. c.

References

Arhar, J. M., Holly, M. L., & Kasten, W. (2001). *Action research for teachers*. Upper Saddle

River, NJ: Merrill Prentice Hall.

Deming, W. E. (1986). *Out of the crisis*. Boston, MA: MIT Press.

DiBella A. J. (2001). *Learning practices: Assessment and action for organizational improvement*. Upper Saddle River, NJ: Prentice-Hall, Inc.

Fraenkel, J. R. & Wallen, N. E. (2006). *How to design and evaluate research in education* (6th ed.). Boston, MA: McGraw-Hill.

McMillan J. H. & Schumacher, S. (2006). *Research in education: Evidence based inquiry* (6th ed.). Boston, MA: Pearson.

Juran, J. M. (1989). *Juran on leadership for quality*. New York, NY: The Free Press.

Kirkpatrick, D. L. & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels* (3rd ed.). San Francisco, CA: Berrett-Koehler Publishers, Inc.

Kirkpatrick, D. L. & Kirkpatrick, J. D. (2007). *Implementing the four levels* (3rd ed.). San Francisco, CA: Berrett-Koehler Publishers, Inc.

Liebowitz, J. (2000). *Building organizational intelligence: A knowledge management primer*. New York, NY: CRC Press.

Rossi, P. H., Lipsey, M. W. & Freeman, H. E. (2004) *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage Publications.

Sagor, R. (1992). *How to conduct collaborative action research*. Alexandria, VA: Association for Supervision and Curriculum Development.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage.

Senge, P. M. (1990). *The fifth discipline: The art & practice of the learning organization*. New York, NY: Doubleday.

Stufflebeam, D. L. (2001). *Evaluation models* (New Directions in Program Evaluation No. 89). San Francisco, CA: Jossey-Bass.

Tyler, R.W. (1949). *Basic principles of curriculum and instruction*. Chicago, IL: University of Chicago.

Yarbrough, D. B., Shulha, L. M. Hopson, R. K. & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.) Thousand Oaks, CA: Sage.