EEN
Evaluation 101
Introduction to Program Evaluation

June 22, 2011

Prepared by:

Yvonne M. Watson
U.S. Environmental Protection Agency

Annelise Carleton-Hug
Trillium Associates

Christina Kakoyannis
National Fish and Wildlife Foundation

Workshop Purpose and Objectives

- Help participants gain familiarity with program evaluation concepts, uses, and methods.
- Provide participants with a framework for evaluating their environmental programs.
- Give participants the opportunity to apply evaluation concepts through interactive exercises.

Workshop Agenda

Module 1: **Program Evaluation: Definition, Uses, Types**

- What is and isn't Program Evaluation?
- What and Why Should We Evaluate?
- Types of Evaluators and Evaluations

Module 2: **Designing the Evaluation**

I. Identify Evaluation Team
II. Develop an Evaluation Plan
III. Model the Program
IV. Develop Evaluation Questions
V. Identify Existing and Needed Data
VI. Identify Performance Measures
VII. Select an Evaluation Design
VIII. Select Data Collection Methods
IX. Analyze & Interpret Data
X. Write & Disseminate Report

Annelise Carleton-Hug

- B.S. Wildlife Management, Ohio State University
- M.S. Fisheries Biology, Humboldt State University
- Ph.D. Human Dimensions of Fisheries & Wildlife, Michigan State University
- Program Evaluator, 2003 –
- American Evaluation Association: Chair, Environmental Program Evaluation Topical Interest Group, 2009-present

www.trilliumassociates.com

Yvonne M. Watson

- B.A. Urban Studies, Wellesley College
- M.A. Urban Planning, UCLA
- Program Evaluator, 2002 –
- Evaluation Support Division, EPA
- American Evaluation Association: Co-Chair, Environmental Program Evaluation Topical Interest Group, 2009-present

watson.yvonne@epa.gov

Christina Kakoyannis

- Ph.D. Forest Resources, Oregon State University
- Program Evaluator, 2001 –
- National Fish and Wildlife Foundation
- American Evaluation Association: Co-Chair, Environmental Program Evaluation Topical Interest Group, 2009-present

christina.kakoyannis@nfwf.org

Orientation Exercise

- The one thing we want to learn…
- The one thing about evaluation that concerns us most now is…

Module 1:

Program Evaluation:

Definitions, Uses, Types

The Evolution of Evaluation

- Origins of Program Evaluation
- A Maturing Profession
- Environmental Program Evaluation

What is Program Evaluation?

Definitions:

- "The systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming." (Patton 1997)

- A systematic study that uses measurement and analysis to answer specific questions about *how well a program is working to achieve its outcomes and why*. (GAO)1Performance Measurement and Program Evaluation: Definitions and Relationships http://www.gao.gov/new.items/11646sp.pdf

Defining Characteristics (Russ-Eft and Preskill 2001)

- Systematic
  - A planned and purposeful activity
- Retrospective
  - Learning from experience
- Analytical
  - Involves collecting data to answer questions
- Informative
  - Enhances knowledge and/or decision-making
- Useful
  - Yields judgment of the merit, worth, or value of something

What *isn't* Program Evaluation?

- Informal judgments
- Collecting information to support a particular agenda
- Auditing
- Basic research
- Performance Measurement: The ongoing monitoring and reporting of program progress and accomplishments, using pre-selected measures.
  - *However,* performance measurement data can provide information needed to conduct the evaluation.
- Logic Model: A diagram and text that illustrates the logical (causal) relationships among program elements and the problem to be solved.
  - *However,* logic models can be used in developing evaluation questions.

Differences between Performance Measurement and Evaluation

Performance Measurement

- Ongoing monitoring and reporting of program performance.
- Tends to focus on achievement of priority program objectives.
- Primarily answers “What?” “How much?” “To what extent?”
- Early warning to management.

Program Evaluation

- In-depth, systematic study conducted periodically.
- Can be used to examine broader range of information on program performance than is feasible to monitor on an on-going basis.
- Explains why the results occurred.
- Longer term review of effectiveness.

What Can Be Evaluated?

- Programs: “A set of planned activities directed towards bringing about specified change(s) in an identified and identifiable audience” (Smith 1989)
  - E.g., H2E, training seminars
- Projects, initiatives, Processes, Systems
  - E.g., process for hiring contractors
- Products
  - E.g., Consumer Reports
- Services
  - E.g., customer service
- People
  - E.g., personnel evaluations

Why Evaluate?

Good Program Management:

- Ensure program goals and objectives are being met.
- Help prioritize resources by identifying the program services yielding the greatest environmental benefit.
- Learn what works well, what does not, and why.
- Learn how the program could be improved.

Federal Laws and Guidance Influencing Evaluation

- Government Performance and Results Modernization Act of 2010:
- President's Executive Order 13450 "Improving Government Program Performance" (2007)
- Office of Management and Budget memo "Increased Emphasis on Program Evaluations" (2009)

Types of Evaluation

Formative

- Purpose: Improve program
- Audience: Program managers and staff
- Diagnostic
- Asks:
  - Where are we?
  - Where do we what to be?
  - What are we doing?
  - What needs to be improved?
  - How can it be improved?
- Primarily internal evaluators

Summative

- Purpose: Decision-making
- Audience: Program managers, potential consumers and funders
- Judgmental
- Asks:
  - What did we do?
  - What did we achieve?
  - Are we effective? How effective?
  - Why or why not effective?
  - At what cost?
- Primarily external evaluators

(Adapted from Worthen et al. 1997.)

Types of Evaluation across the Life of Program

Design (Developmental) Evaluation

- When: During initial stages of program development & design
- Needs assessments
- Typical Questions: *Where are we now? Where do we want to be? What are appropriate goals and objectives?*

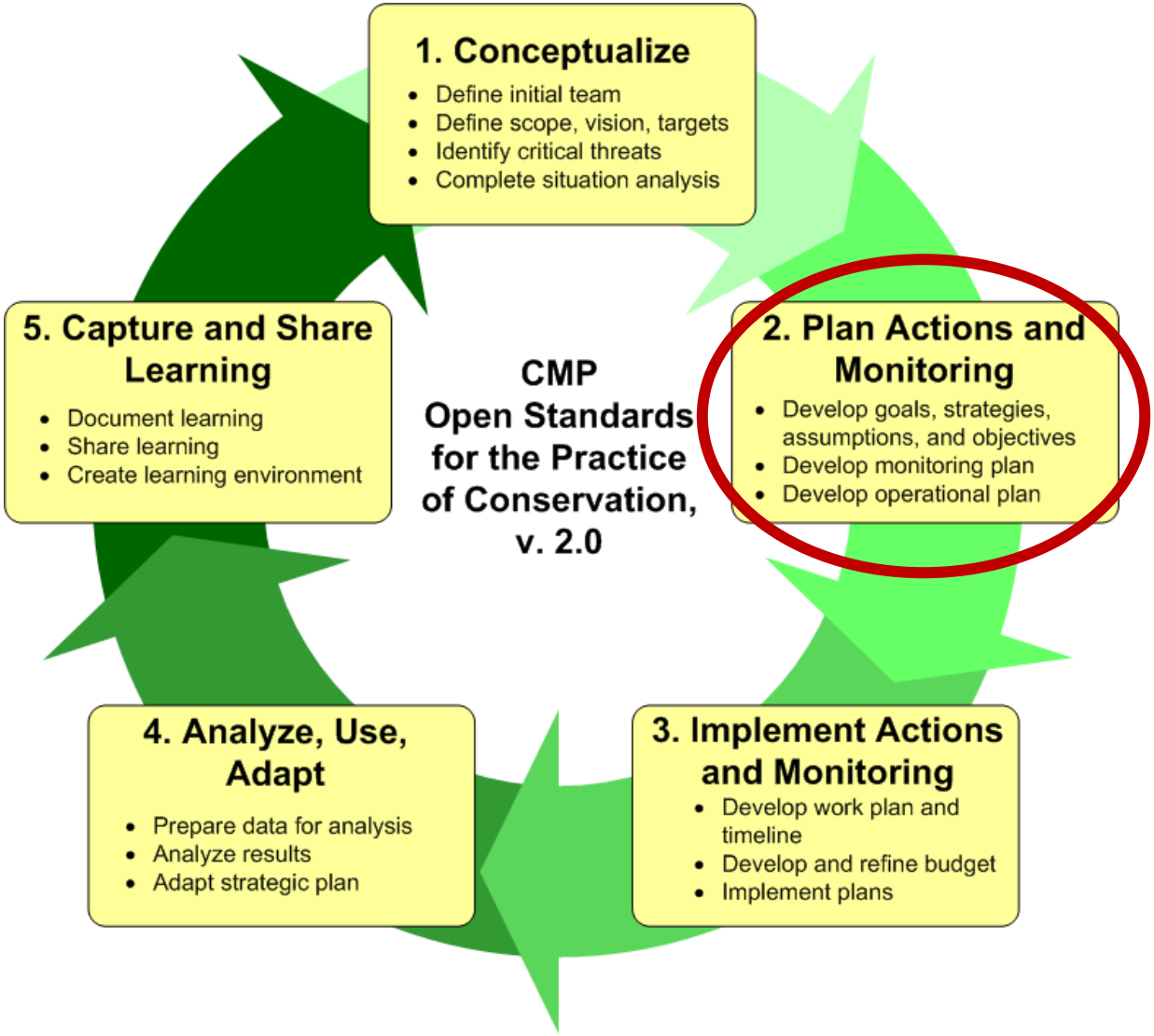
Process (Implementation) Evaluation

- When: During existing program implementation
- Focused on assessing process of program implementation
- Typical Questions: *To what degree is the program strategy being implemented as intended? Are our processes and systems appropriate given our objectives?*

Outcome and Impact Evaluation

- When: Retrospective on mature program
- Examines program's short, intermediate, & long-term outcomes.
- Typical Questions: *Were outcomes achieved? How can we improve our level of effectiveness? Unanticipated results?*

Steps in Managing a Program or Project Adaptively



Guiding Principles for Evaluators (American Evaluation Association, 1995)

- **Systematic Inquiry** - Evaluators conduct systematic, data-based inquiries about what is being evaluated.
- **Competence** - Evaluators provide competent performance to stakeholders.
- **Integrity/Honesty** - Evaluators ensure the honesty and integrity of the entire evaluation process.
- **Respect for People** - Evaluators respect the security, dignity, and self-worth of the respondents, program participants, clients, and other stakeholders with whom they interact.
- **Responsibilities for General & Public Welfare** - Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare.

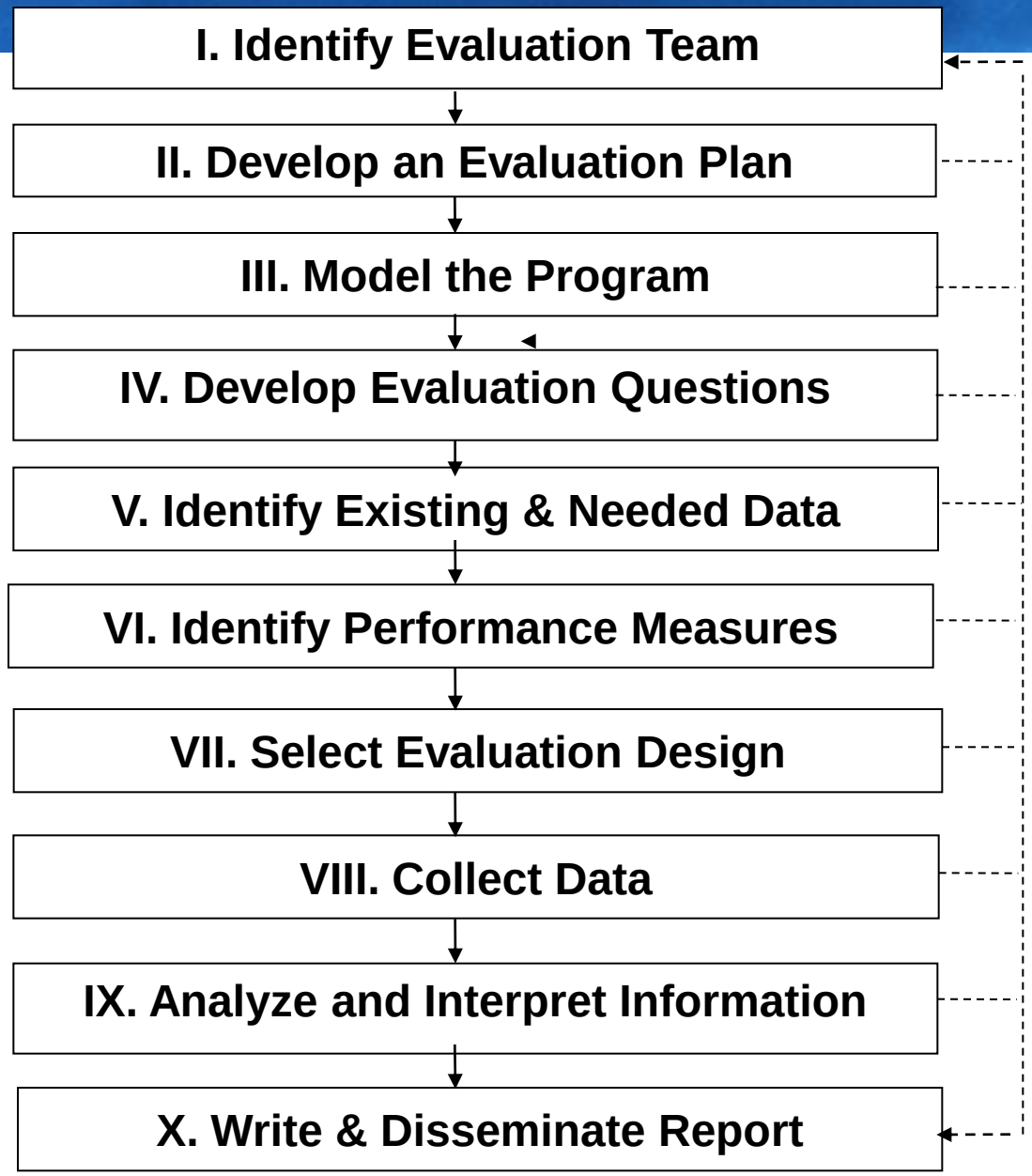
Standards for Evaluation - Joint Committee on Standards for Educational Evaluation (1994)

- **Utility** - standards intended to ensure that an evaluation will serve the practical information needs of the intended users.
- **Feasibility** - standards intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.
- **Propriety** - standards intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.
- **Accuracy** - standards intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.

Module 2:

Designing the Evaluation

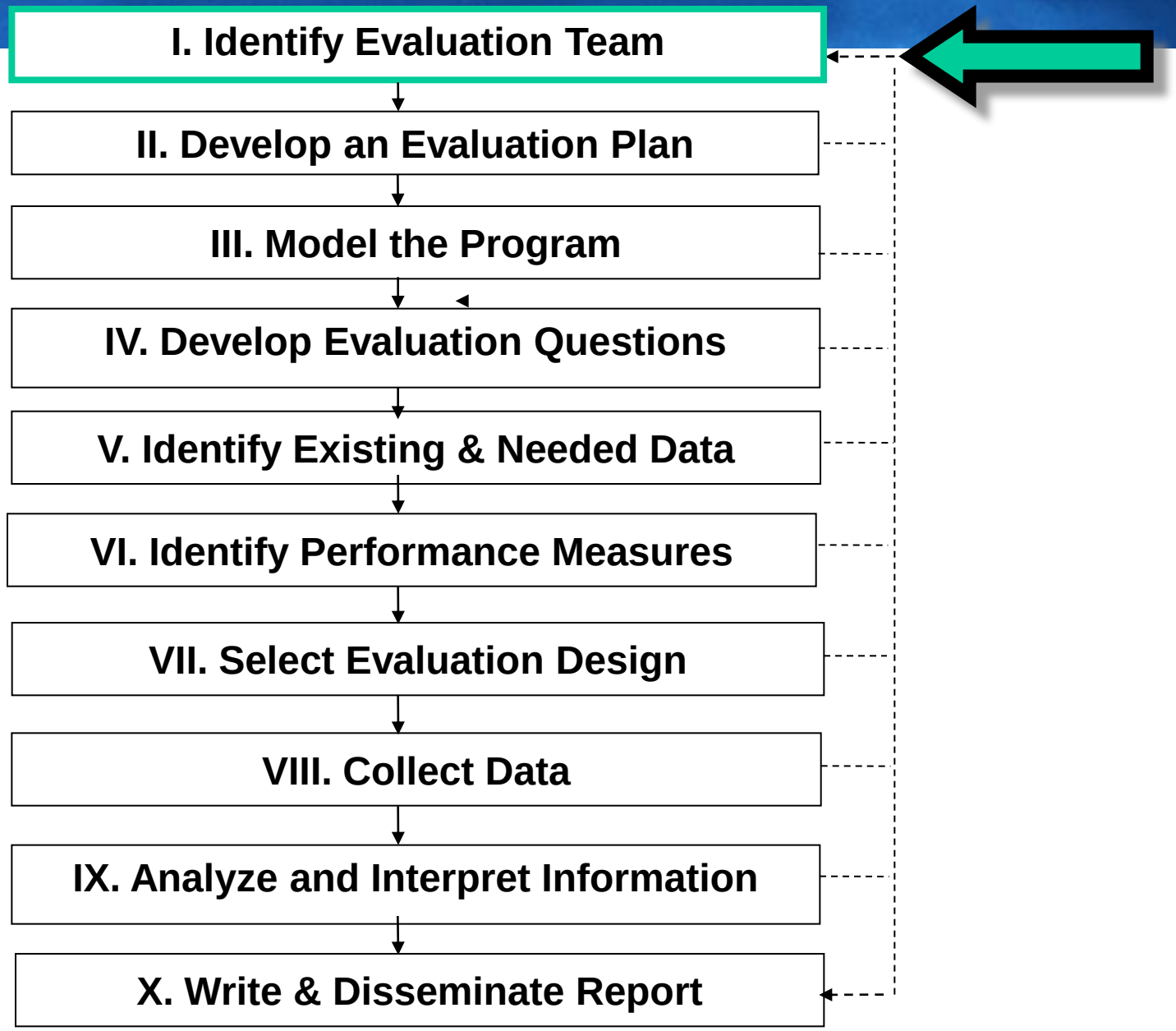
Steps to Completing an Evaluation



From Theory/Concepts to Practice

An Evaluation of the Yellowstone National Park's Electronic Field Trip

Steps to Completing an Evaluation

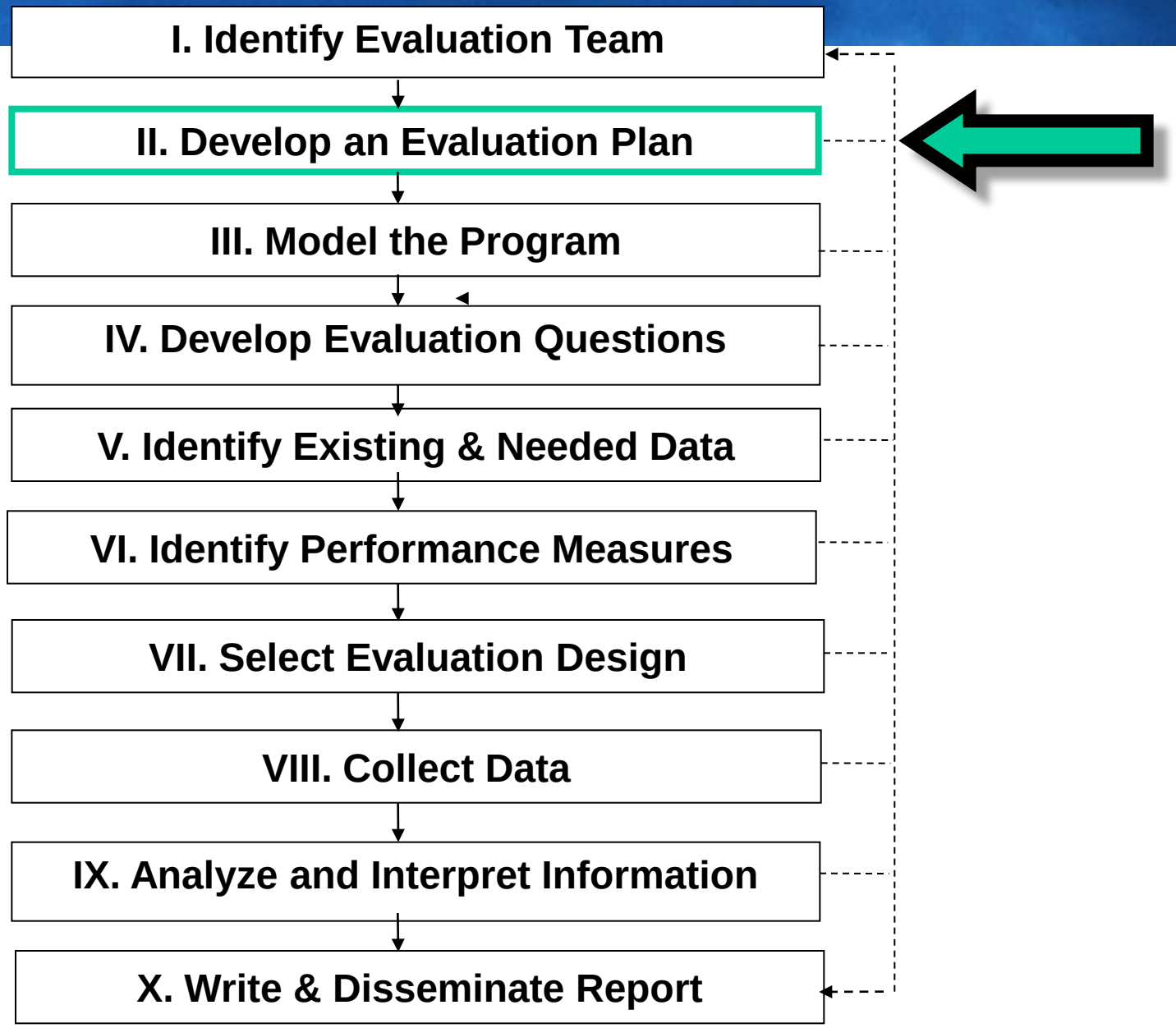


Identify Evaluation Team Members

Select diverse team members:

- Individuals responsible for designing, collecting, and reporting information used in the evaluation
- Individuals with knowledge of the program
- Individuals with a vested interest in the conduct/impact of the program
- Individuals with knowledge of evaluation
- Identify a Skeptic!

Steps to Completing an Evaluation



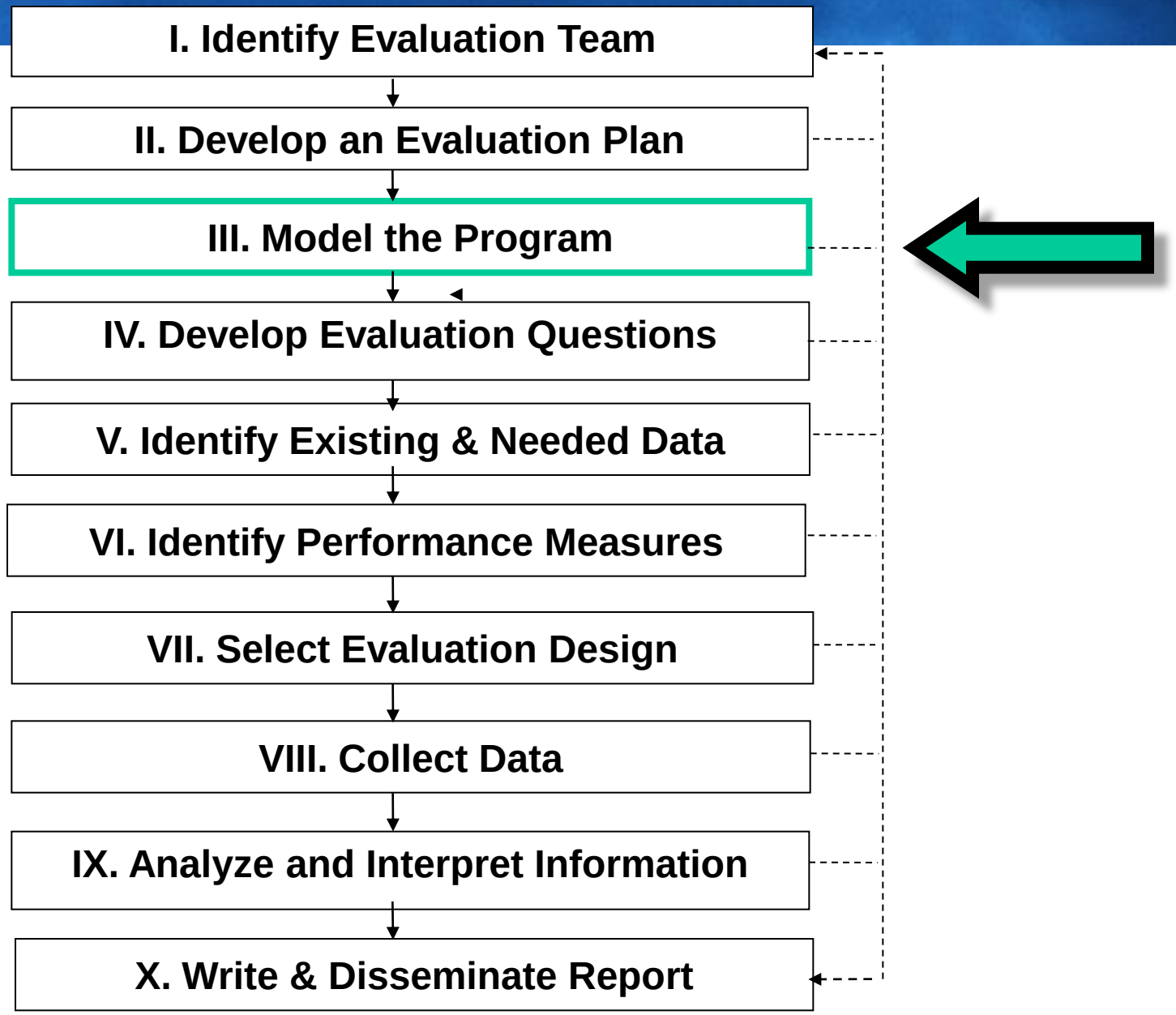
The Evaluation Plan

- **What:** Brief document describing evaluation purpose, audience, scope, design, & methods.
- **Why:** The purpose is to clearly articulate and communicate expectations for the evaluation.
- **Who:** Developed by one or more team members based on team's common understanding.
- **When:** Can be developed at any point from initial selection of the program through development of the research design.
  - Plans are living documents and need to be revised to account for changes in evaluation objectives or methods.

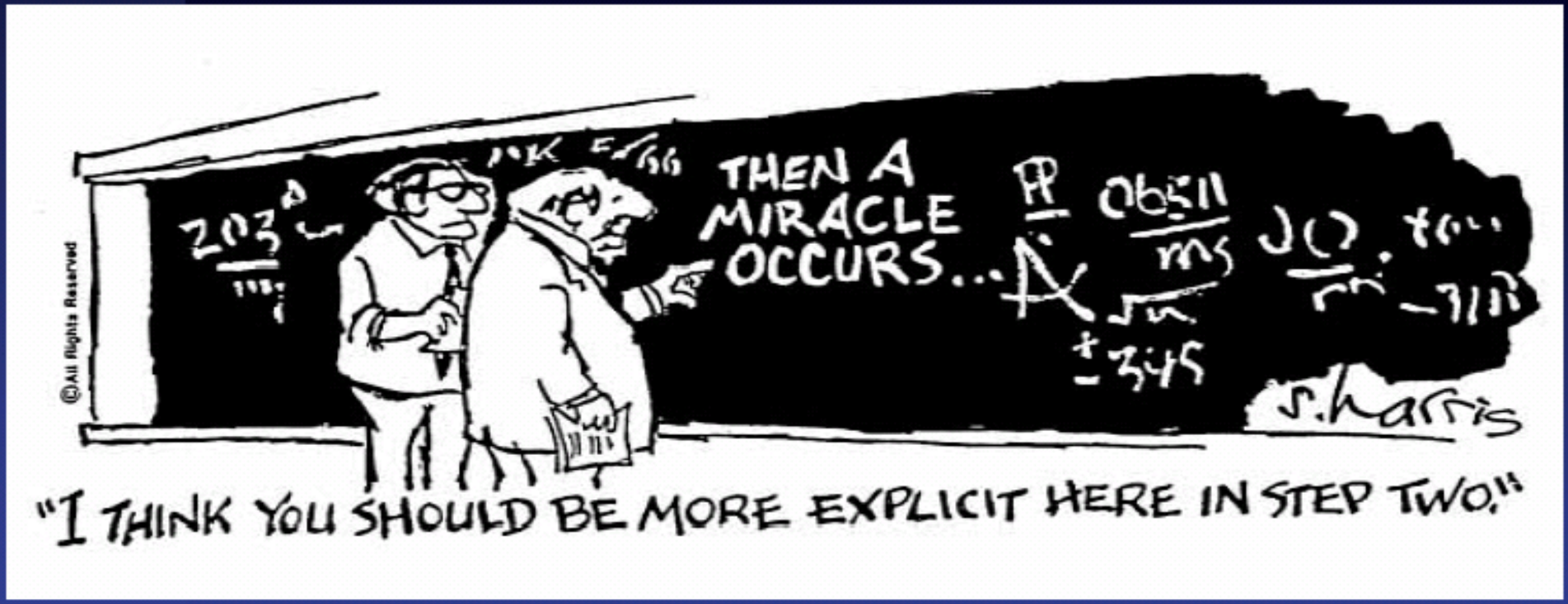
Components of an Evaluation Plan

- Purpose of the evaluation/ Evaluation questions
- Primary audience
- Context (organizational, management, political)
- Data collection methods and analysis
- Evaluation design
- How evaluation findings will be reported
  - Consider different formats for different target audiences
- Expectations for roles and communication among evaluators, program staff/managers, and key stakeholders
- Resources available for evaluation (staff, budget)
- Timeline for evaluation
  - Note: Save sufficient time to develop evaluation questions and analyze data thoroughly.

Steps to Completing an Evaluation



The Logic Model



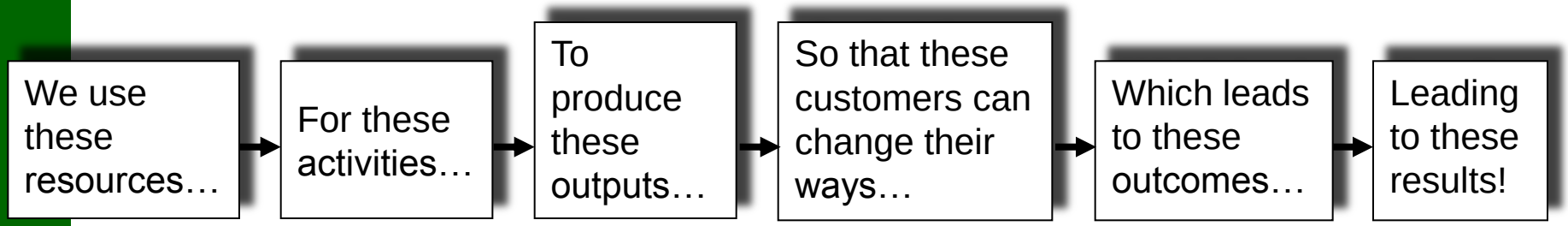
Model the Program

- Model the program
  - Create a diagram that illustrates how the program is supposed to work

- Similar Concepts
  - Program theory
  - Program roadmap
  - Theory of change
  - Program hypothesis
  - Results chain

What is a Logic Model?

A logic model is a diagram and text that describes/ illustrates the logical (causal) relationships among program elements and the problem to be solved, thus defining measurements of success.

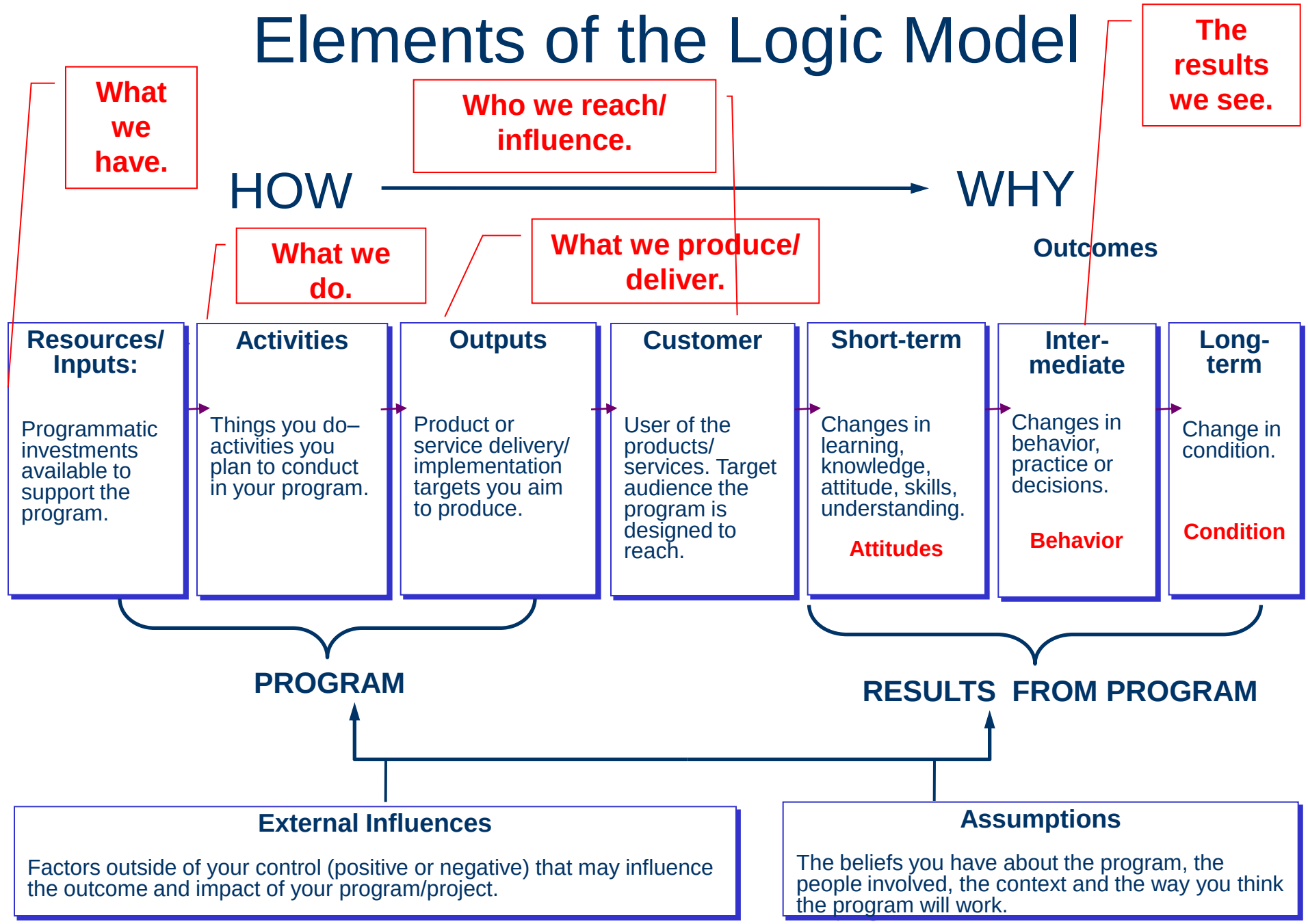


Benefits of Developing Logic Models

If you don't know where you are going, you might wind up someplace else. - Yogi Berra

- Clarifies program objectives
- Defines anticipated results chains
- Establishes framework for evaluation
- Process can improve program “cohesiveness”

Elements of the Logic Model



Steps in the Logic Model Process

1. Define the problem and context for the program or project and determine what aspect of your program/project you will model.
2. Clarify the program goal and define the elements of the program in a table.
3. Verify the elements of the table with stakeholders.
4. Develop a model describing logical relationships.
5. Verify the Logic Model with stakeholders.

Then use the Logic Model to identify and confirm performance measures and in planning and evaluation.

Step 2. Clarify the program goal and define the elements of the program or project in a table

WHAT and WHY

- HOW -			WHO	Outcomes		
Resources/ Inputs	Activities	Outputs	Customers reached	Short-term (change in attitude)	Intermediate (Change in behavior)	Long-term (change in condition)

External Influences:

Step 3. Verify the logic table with stakeholders

				Outcomes		
Resources/ Inputs	**Activities**	**Outputs**	**Customers Reached**	**Short-term** (Change in Attitude)	**Intermediate** (Change in Behavior)	**Long-term** (Change in Condition)
$62.5M 65 FTE (AIEO, Regions, National Program Managers) Agency Technical Expertise	Provide funding to Tribes Provide technical assistance to Tribes	GAP Grants Technical and media-specific training	Tribal Executives Inter-Tribal Consorita Executives Tribal environmental employees funded by GAP	Increase understanding of the process required for an environmental program	Legal capability Enforcement capability Technical capability Communications capability Administrative capability	Established capacity to plan, develop, implement & manage environment-al programs Improved environment-al conditions in Indian Country

External Factors: Tribal leadership, vision, continuity, priority of environmental issues, education levels, staff turnover, and resource levels.

Step 4. Develop a diagram and text describing logical relationships

- Draw arrows to indicate/link the causal relationships between the logic model elements.

- Limit the number of arrows. Show only the most critical feedback loops.
- Work from both directions (right-to-left and left-to-right):
  - Ask "*How-Why*" questions:
    - Start with Outcomes and ask "How?"
    - Start at Activities and ask "Why?"
  - Ask "*If-Then*" questions:
    - Start at Activities and move along to Outcomes asking, "*If this, then that*?"

Two Important Rules to Follow

- For every *action* identified in the Logic Model, the must be an *output that connects* to an *outcome* through a *specific customer.*

OR

- An *action* must produce an *output* that becomes a *key input* to another *activity*.

THINK CONNECTIONS!

Key Questions to Consider…

- Are the program's outcomes described?
- Are the program's customers described?
- Are the program's major resources, activities and outputs described and do they make sense?
- Are there factors/issues that might influence the program's ability to achieve its goal?

Tribal General Assistance Program

U.S. EPA Tribal General Assistance Program (GAP)

The primary purpose of GAP is to help federally recognized tribes and intertribal consortia build the basic components of a tribal environmental program, which may include planning, developing, and establishing the administrative, technical, legal, enforcement, communication, and outreach infrastructure.

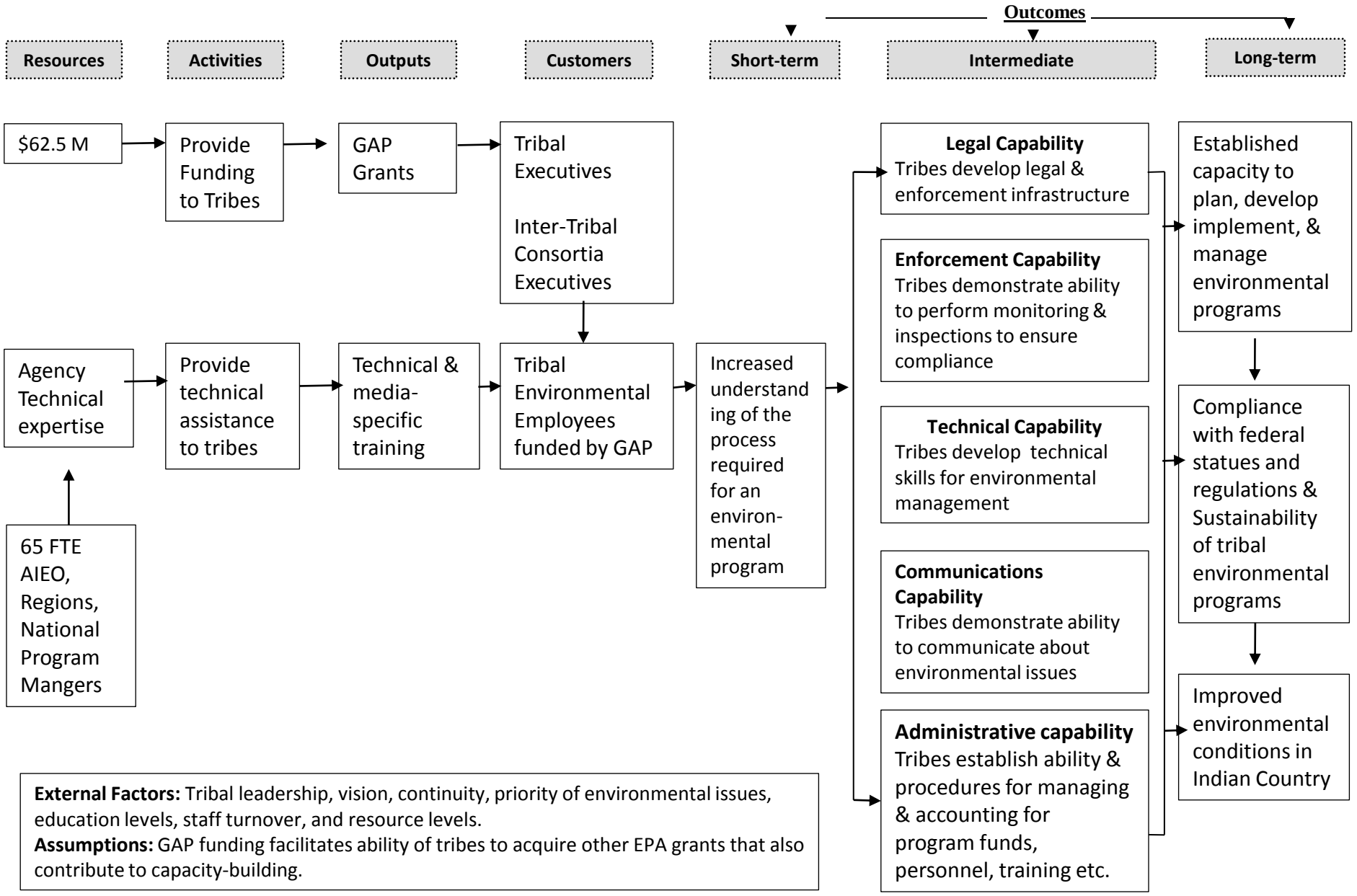
Evaluation Purpose:

Determine how effective GAP has been in building tribal environmental capacity with those tribes receiving funds.

Report Available at:
http://www.epa.gov/evaluate/GAPFinalReport.pdf

General Assistance Program (GAP) Grant Program Logic Model

Program Goal: The primary purpose of GAP is to help federally recognized tribes and intertribal consortia build the basic components of a tribal environmental program, which may include planning, developing, and establishing the administrative, technical, legal, enforcement, communication, and outreach infrastructure.

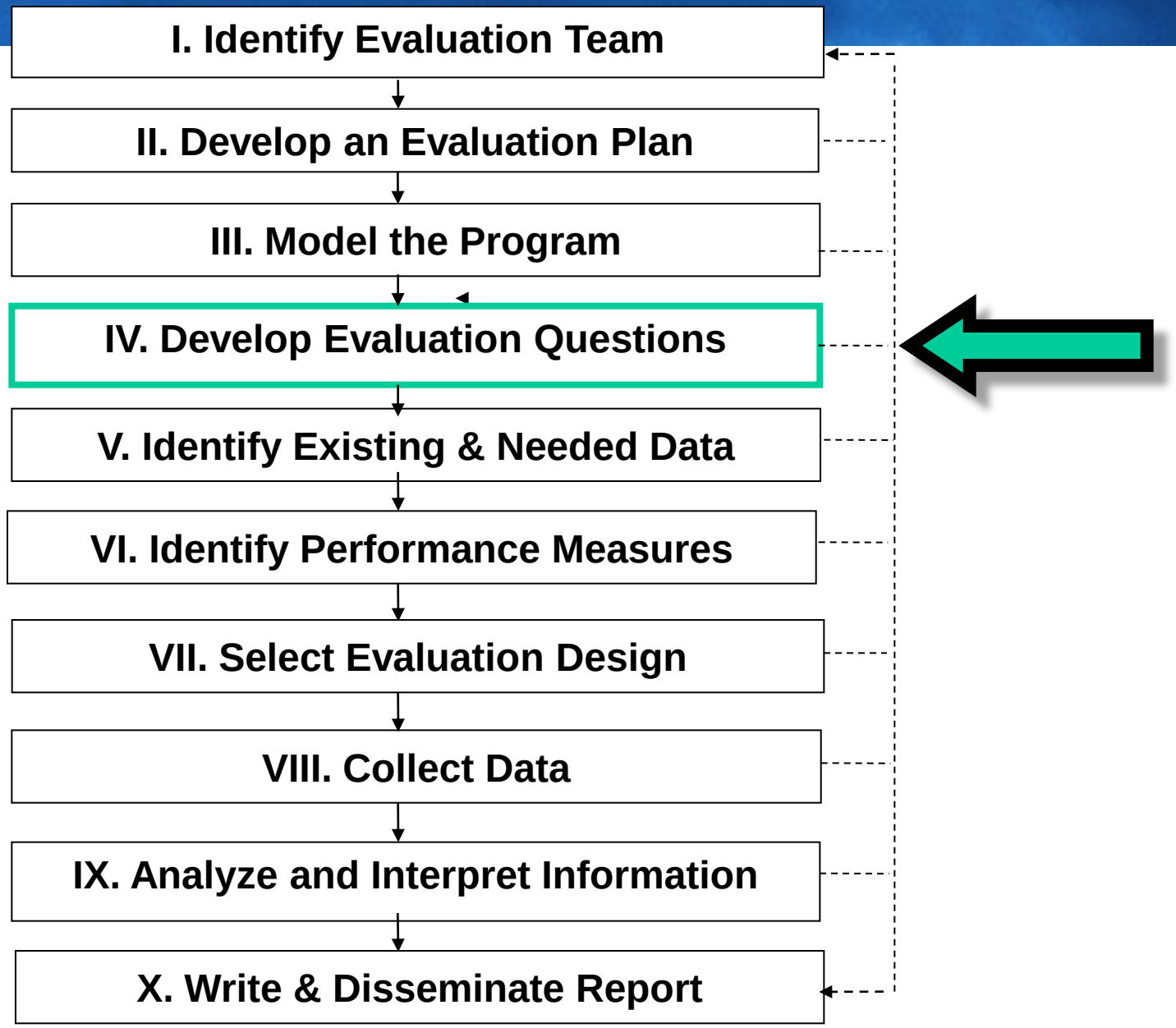


External Factors: Tribal leadership, vision, continuity, priority of environmental issues, education levels, staff turnover, and resource levels.
Assumptions: GAP funding facilitates ability of tribes to acquire other EPA grants that also contribute to capacity-building.

Exercise 1:

Develop a Logic Model

Steps to Completing an Evaluation



What are Evaluation Questions?

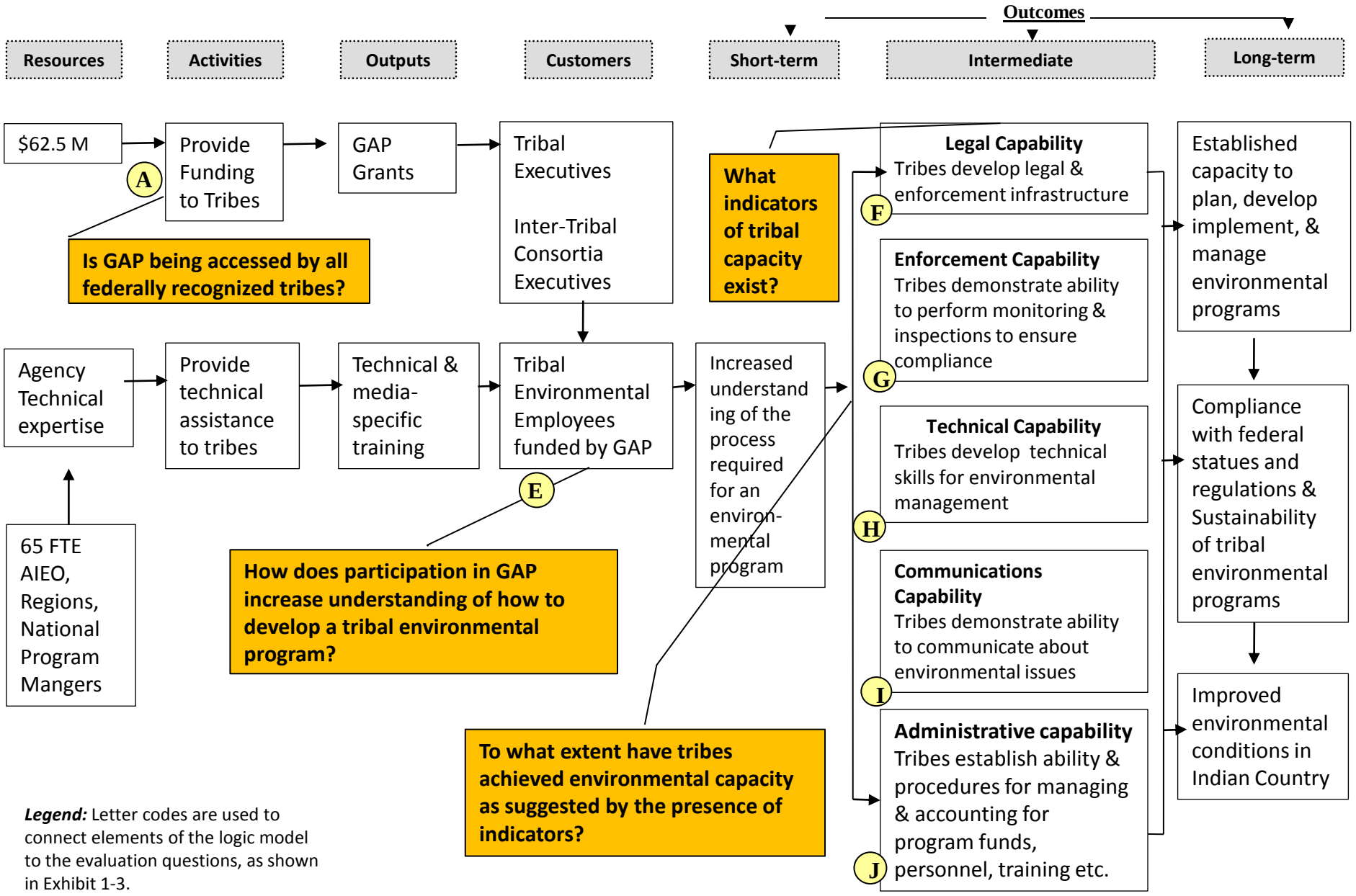
- Questions (at any point on the performance spectrum/ logic model) that the evaluation is designed to answer.
- They should reflect stakeholders' needs.
- Evaluation questions are **KEY** because they:
  - Frame the scope of the evaluation
  - Drive the evaluation design, data collection, and reporting

Steps for Developing Questions

1. Use program's logic model to:
   a. Review program purpose and objectives.
   b. Identify what aspects of program you want to evaluate.
2. Brainstorm a variety of different questions by asking:
   a. I would really like to know _____ about this program.
   b. What information about this program, if it were available, would make a difference in what I do?
3. Obtain feedback and suggestions from stakeholders.
4. Group questions by type (e.g., design, process, outcome evaluation *or* by logic model categories).
5. Select highest priority questions for the evaluation.

General Assistance Program (GAP) Grant Program Logic Model

Program Goal: The primary purpose of GAP is to help federally recognized tribes and intertribal consortia build the basic components of a tribal environmental program, which may include planning, developing, and establishing the administrative, technical, legal, enforcement, communication, and outreach infrastructure.



Legend: Letter codes are used to connect elements of the logic model to the evaluation questions, as shown in Exhibit 1-3.

Tribal GAP Evaluation Questions and the Logic Model

Exhibit 1-3:
Relationship Between Evaluation Questions and Logic Model

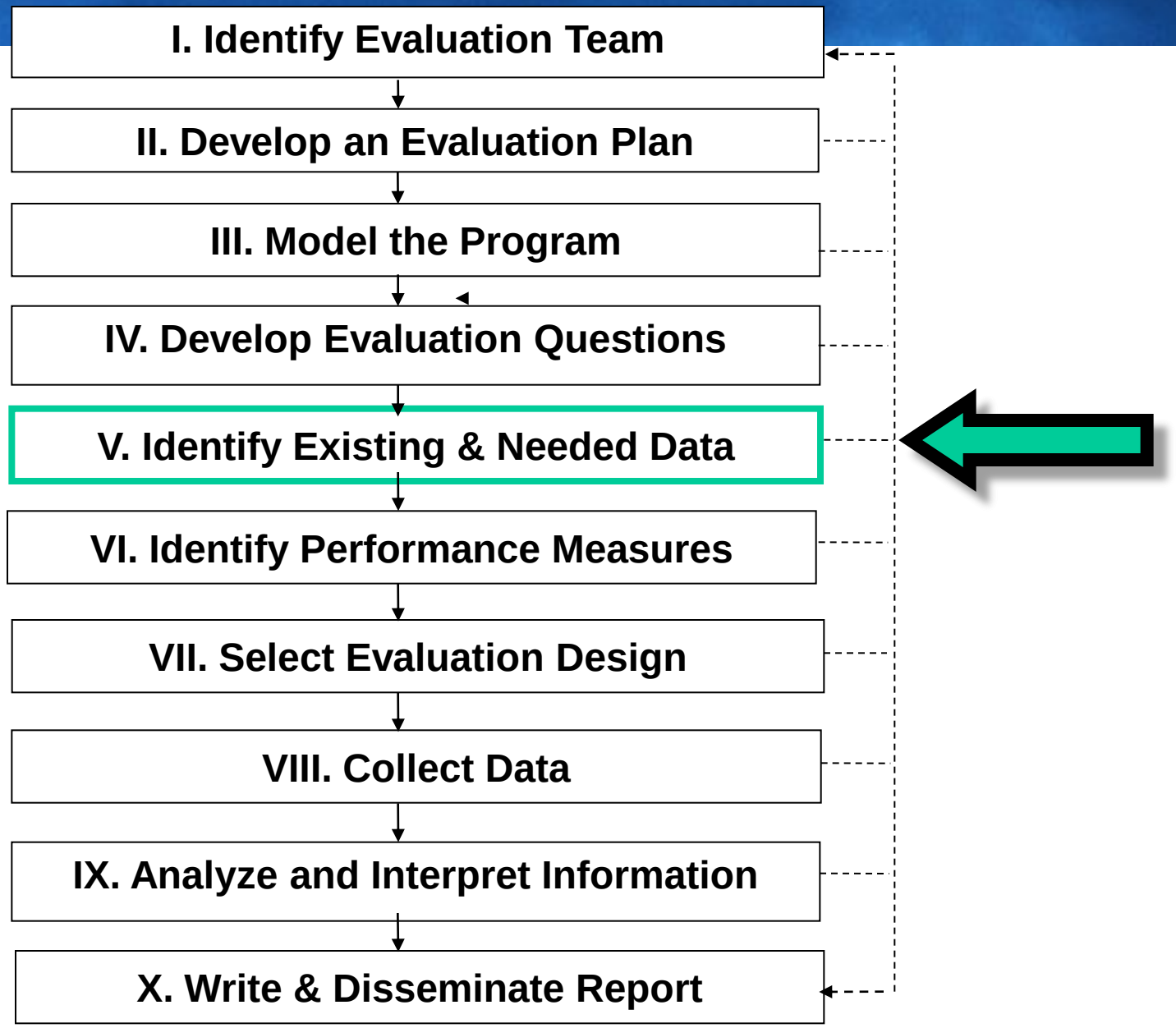
Evaluation Question	Component of the Logic Model
1a. Is GAP being accessed by all federally recognized tribes?	A
1b. Why are some tribes not involved in GAP?	A
1c. Are there tribes that received GAP grants at one time but which no longer receive GAP grants? If so, why?	A
2a. Are tribal governments using the resources (technical, fiscal, and programmatic) provided as a component of GAP? How often are GAP resources accessed?	A B C
2b. How are tribes using GAP resources?	D F G H I J
2c. To what extent have tribes met program expectations for grants management, execution of administrative functions, and carrying out proposed activities?	J
2d. How does participation in GAP increase understanding of how to develop a tribal environmental program?	E

Tribal GAP Evaluation Questions and the Logic Model

Exhibit 1-3:
Relationship Between Evaluation Questions and Logic Model

Evaluation Question	Component of the Logic Model
3a. What indicators of tribal environmental capacity exist? 3b. To what extent have tribes achieved environmental capacity as suggested by the presence of these indicators?	F G H I J K
3c. What factors contribute to the achievement of environmental capacity, and what is the impact of each factor?	
Tribal Priorities Tribal	L
Tribal Staffing	D
Tribal Funding	A L
Communication	K
Regional Activities	M
4. Is the GAP providing adequate outputs to achieve tribal goals and priorities?	*Not directly shown in logic model*
5. To what degree does GAP support EPA's strategic goal of increasing tribes' ability to build environmental program capacity?	N

Steps to Completing an Evaluation



Identify Existing & Needed Data

- Purpose: To learn what data the program needs to compile or collect to answer the evaluation question(s).
- Evaluation team should ask:

1. What information (data) do we need to answer the evaluation questions?
2. Do we have existing performance measures or data collection efforts collecting similar information?
   - If yes, is this information sufficient for answering the evaluation questions?
3. What additional information do we need to collect to answer the evaluation questions?
   - What performance measure(s), if any, should we develop to obtain this information?

Tribal GAP: Needed Information

Evaluation Question	Information That Can Help Answer Question
1a. Is GAP being accessed by all federally recognized tribes?	Number and percentage of federally recognized tribes that have ever received GAP funds
1b. Why are some tribes not involved in GAP?	Regional coordinators' perceptions as to why tribes may not seek GAP funding (these may include having access to other sources of funds, the perception that participation in GAP is too onerous, etc.)
1c. Are there tribes that received GAP grants at one time but which no longer receive GAP grants? If so, why?	Regional coordinators' perceptions as to why tribes may have dropped of GAP grant rolls.
2a. Are tribal governments using the resources (technical, fiscal, and programmatic) provided as a component of GAP? How often are GAP resources accessed?	▪Number and type of GAP resources that have been delivered to tribes: ▪GAP funding provided to tribes (i.e., fiscal resources) ▪Technical assistance and media specific trainings (i.e., technical resources) ▪Grants management training (i.e., programmatic resources) ▪Tribal access of GAP resources: ▪Tribal participation in technical and media-specific trainings ▪Tribal participation in grants management training
2b. How are tribes using GAP resources?	Tribal staff and activities funded through GAP (*include solid waste implementation*)
2c. To what extent have tribes met program expectations for grants management, execution of administrative functions, and carrying out proposed activities?	▪Regional coordinators' perceptions about the quality, timeliness, and completeness of work plans and progress reports received ▪Timing of grant end date vs. final close out of the grant - this is an indicator of the degree to which the grantee met program expectations - the shorter the period of time between grant end date and final closeout, the more likely that tribes met expectations. Results of administrative post award monitoring audits.

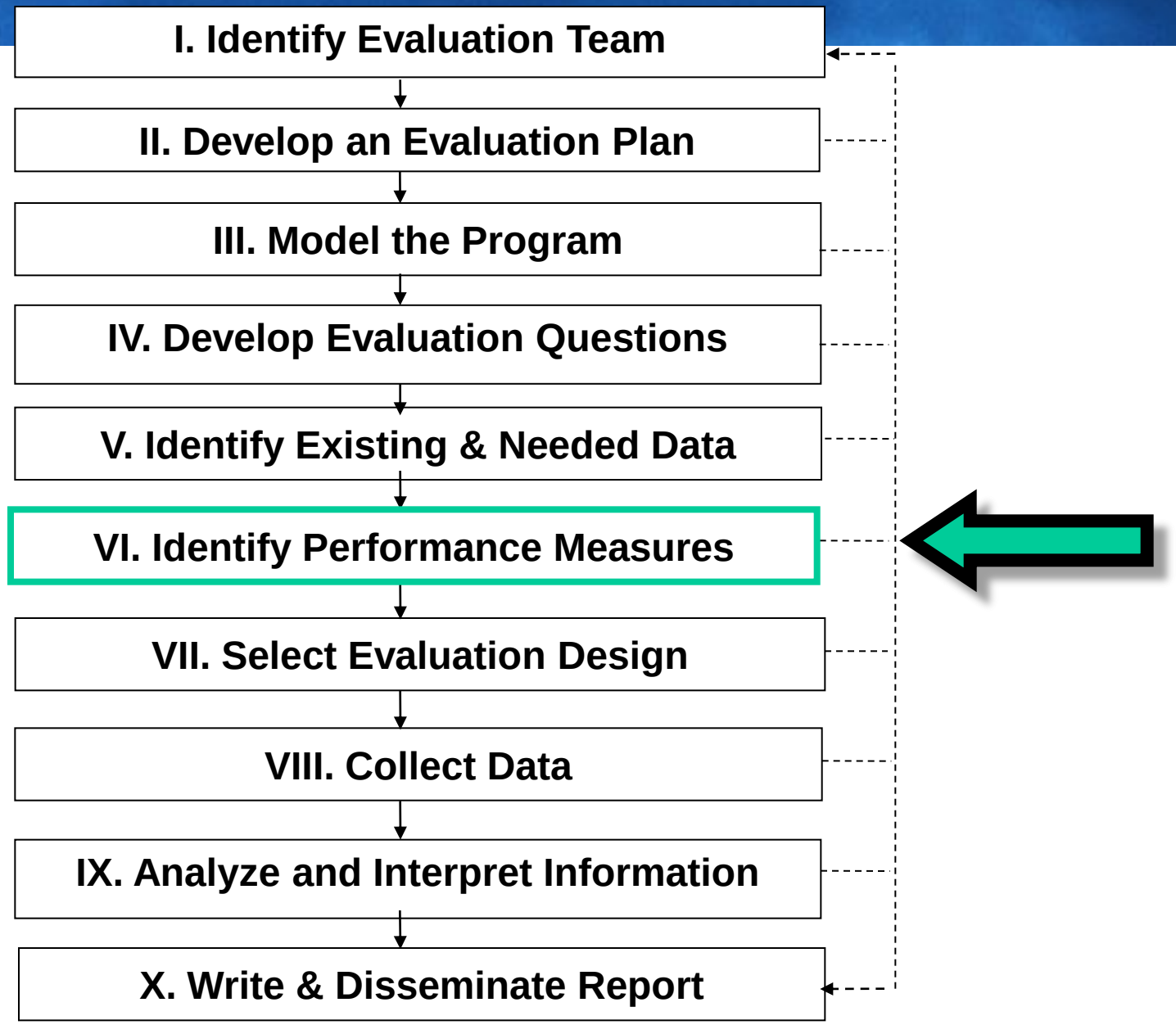
Tribal GAP: Needed Information

Evaluation Question	Information That Can Help Answer Question
2d. How does participation in GAP increase understanding of how to develop a tribal environmental program?	▪Self reported increase in knowledge and understanding about the necessary steps in developing a tribal environmental program ▪Self-reported increase in skills needed to develop tribal environmental programs ▪Self-reported change in awareness and commitment to environmental programs in tribes
3a. What indicators of tribal environmental capacity exist? 3b. To what extent have tribes achieved environmental capacity as suggested by the presence of these indicators?	**Overarching Indicator of Tribal Environmental Capacity:** ▪Number of GAP recipients that secured ongoing funding from other EPA sources. [*Note that availability of other sources of funding may be a limiting factor unrelated tribes' environmental capacity*.] **Legal Capability:** Number of GAP recipients that have developed tribal codes, standards, and/or enforcement programs to control pollution **Enforcement Capability:** Presence of tribal environmental staff person(s) charged with enforcement duties **Technical Capability :** Number of GAP recipients with one or more staff specifically taskedwith managing environmental programs (e.g., Environmental Director) Size and composition of tribal environmental staff Number of environmental programs being carried out in different media annually by tribes. Number of GAP recipients that have taken environmental training

Exercise 2:

Develop Evaluation Questions and Identify Existing and Needed Data

Steps to Completing an Evaluation



Why develop performance measures?

Performance Measure: A metric used to gauge program or project performance (AKA):

- Metrics
- Indicators

- Performance measures:
  - Assess the effect of your program
  - help answer the evaluation questions
  - provide a consistent, quantitative way to assess program status/outcomes
  - are paired with qualitative information to provide a comprehensive response to evaluation questions

Example Performance Measures

Element	Definition	Example Measure
Resources/ Inputs	Measure of resources consumed by the organization.	Amount of funds, # of FTE, materials, equipment, supplies (etc.).
Activities	Measure of work performed that directly produces the core products and services.	# of training classes offered as designed; Hours of technical assistance training for staff.
Outputs	Measure of products and services provided as a direct result of program activities.	# of technical assistance requests responded to; # of compliance workbooks developed/delivered.
Customer Reached	Measure of target population receiving outputs.	% of target population trained; # of target population receiving technical assistance.
Customer Satisfaction	Measure of satisfaction with outputs.	% of customers dissatisfied with training; % of customers “very satisfied” with assistance received.
Outcomes	Accomplishment of program goals and objectives (short-term and intermediate outcomes, long-term outcomes--impacts).	% increase in industry’s understanding of regulatory recycling exclusion; # of sectors that adopt regulatory recycling exclusion; % increase in materials recycled.

Example performance measures, cont.

Category	Definition	Examples
Efficiency	Measure that relates outputs to costs.	Cost per workbook produced; cost per inspection conducted.
Productivity	Measure of the rate of production per some specific unit of resource (e.g., staff or employee). The focus is on labor productivity.	Number of enforcement cases investigated per inspector.
Cost Effectiveness	Measure that relates outcomes to costs.	Cost per pounds of pollutants reduced; cost per mile of beach cleaned.
Service Quality	Measure of the quality of products and services produced.	Percent of technical assistance requests responded to within one week.

Key Steps in Identifying Potential Measures

1. Identify the information needed and the audience
   - Review the evaluation questions developed earlier and consider what information is needed to answer these questions (**Who needs to know what about the program, why, and in what format?)**
2. Identify measures in existing documents
   - Previous evaluations, research reports & other sources
3. Identify measures from the Logic Model
   - What aspects of performance are most important to measure (resources, activities, outputs, outcomes)?
   - What contextual factors could influence the program either positively or negatively?
   - Express the element as a measure (i.e., identify how the measure will be calculated, percentages, raw numbers, averages, rates, ratios)

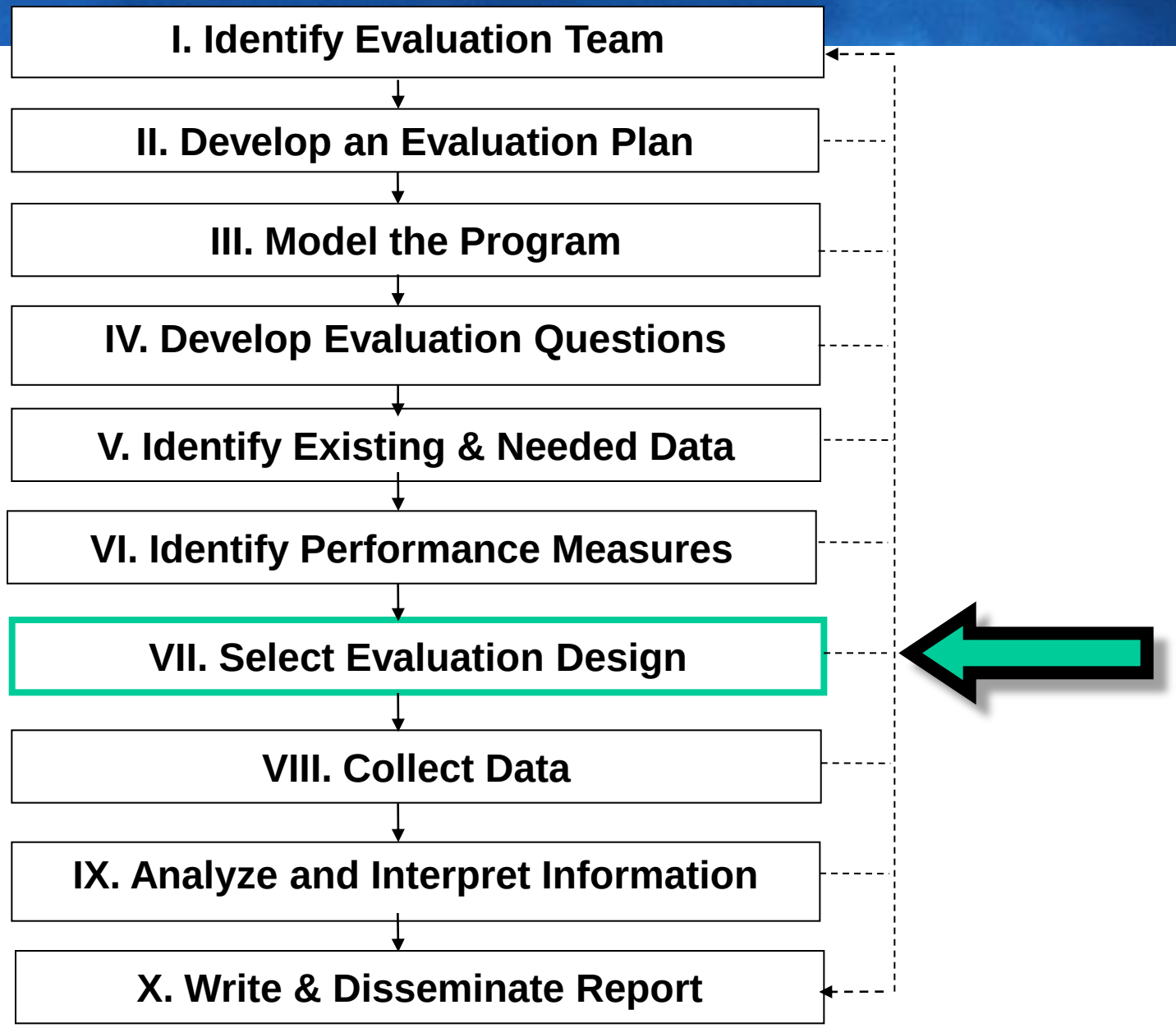
Tribal GAP Performance Measures

	Resources	Activities	Outputs	Customer reached	Short-term Outcome	Intermediate Outcome	Long-term Outcomes
Logic Model Elements	▪$62.5 M ▪65 FTE AIEO, Regions, National Program Managers	▪Provide funding to Tribes ▪Provide technical assistance to Tribes	▪GAP Grants ▪Technical & media-specific training	▪Tribal Executives ▪Inter-Tribal Consortia Executives ▪Tribal Environmental employees funded by GAP	▪Increased understanding of the process required for an environmental program	▪Legal capability ▪Enforcement Capability ▪Technical Capability ▪Communications Capability ▪Administrative Capability	▪Established capacity to plan, develop, implement & manage environmental programs ▪Improved environmental conditions in Indian Country
Example Measures	▪Total GAP funding per year	▪Number and percent of Federally recognized Tribes that have received GAP funds between 1994 and 2004	▪Percentage of tribes that have received technical assistance and media specific training	▪Total number of environmental employees supported with GAP funds ▪Turnover rate for Tribal environmental employees	▪Self reported increase in knowledge and understanding about the necessary steps in developing a tribal environmental program ▪ Self-reported increase in skills needed to develop tribal environmental programs ▪ Self-reported change in awareness and commitment to environmental programs in tribes	▪# of GAP recipients that have developed tribal codes, standards or enforcement programs to control pollution	

Exercise 3:

Identify Performance Measures

Steps to Completing an Evaluation



Evaluation Design

Evaluation Design: The methodological approach used to answer the evaluation questions. A strong design decreases vulnerability to methodological criticism.

Ways to Strengthen the Evaluation Design:

1. **Pre-testing**: Compare outcomes of program participants from both before (i.e., baseline) and after the program intervention.

2. **Control Group**: Compare outcomes of program participants to outcomes of control groups that do not receive program services.

3. **Randomization**: Randomly assign subjects into the treatment (i.e., program) or control groups.

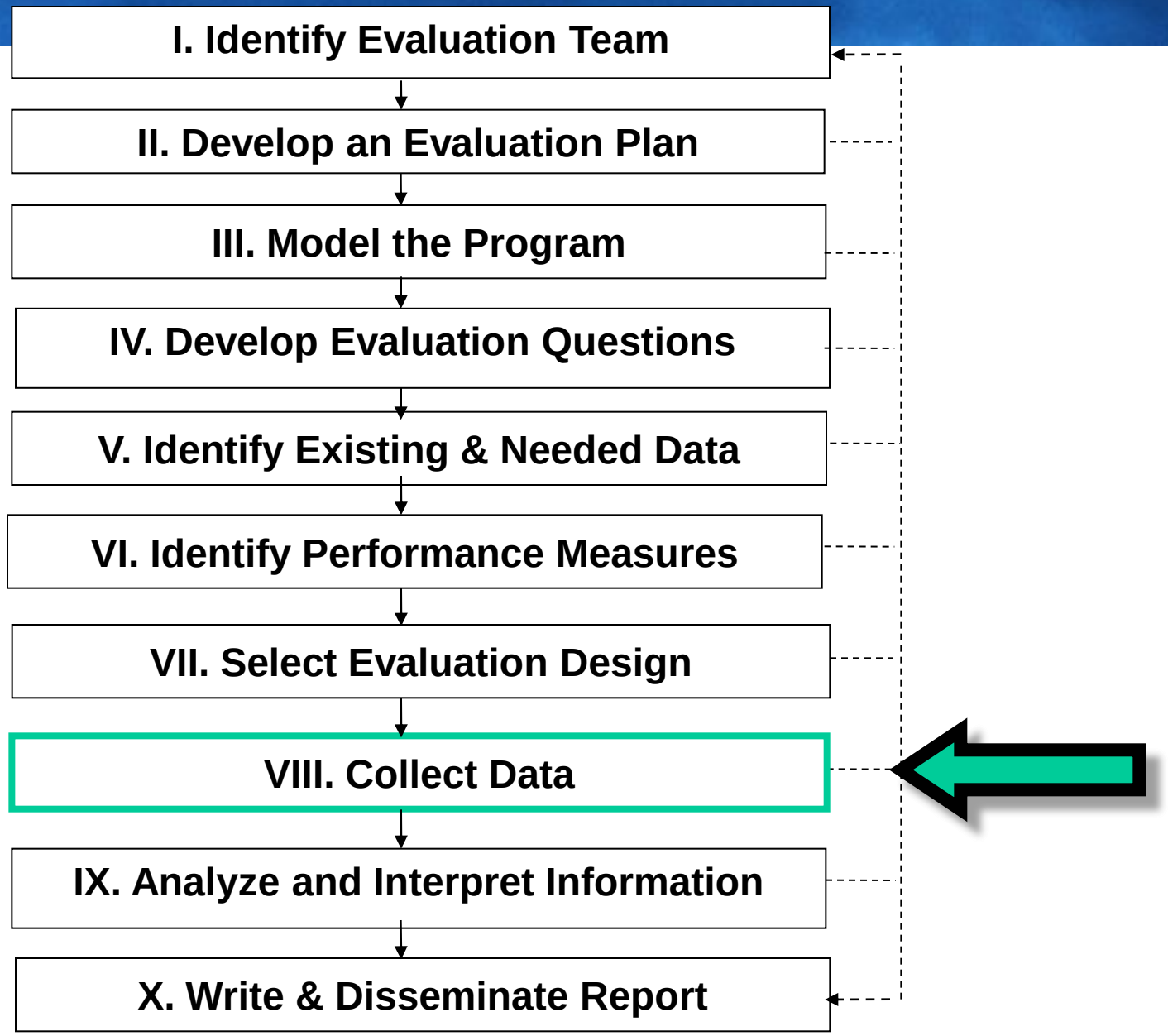
Evaluation 101 Training: Strength of Potential Designs

Weakest

Strongest

1. Administer a test of evaluation concepts to participants who complete the Evaluation 101 workshop.
2. Administer a test of evaluation concepts to participants both before *and* after the workshop. Compare the difference in scores.
3. Administer a test of evaluation concepts to workshop participants both before and after the workshop AND compare the difference in scores with those from a group of similar people who did not attend the workshop.
4. Randomly assign people into the group that will attend the workshop and the group that will not. Administer the tests both before and after the workshop & compare the difference in scores.

Steps to Completing an Evaluation



Considerations in Selecting a Data Collection Method

- Your evaluation question
- Stakeholders' desired sources of data
- Resources (e.g., available skills)
- Time available to collect data
- Data availability
- Validity of different sources of data
- Information Collection Requests (ICR)

Types of Data Collection Methods

- Qualitative Methods:
  - Examples: Interviews, focus groups, direct observation, document review.
  - Often used to obtain information on processes, meanings, in-depth understanding.
- Quantitative Methods:
  - Examples: Survey questionnaires, tests, checklists, monitoring data.
  - Often used to obtain information on outcomes and causal relationships.
- Not a question of either/or, but when to use a method given evaluation question. Using both methods can yield strongest conclusions.

Data Collection Methods

Method	Overall Purpose	Advantages	Challenges
Interviews	To fully understand someone's impressions or experiences, or learn more about their answers to questionnaires	- Get full range and depth of information - Get targeted information - Develops relationship with client - Can be flexible with client	-Time consuming/ costly - Can be hard to compare responses - Interviewer can bias client's responses - Inaccurate recall
Focus Groups	To explore a topic in depth through group discussion, e.g., about reactions to an experience, understanding common complaints	- Quickly and reliably get common impressions - Can be efficient way to get much range and depth of information in short time	Can be hard to compare responses - Need good facilitator for safety and closure - Difficult to schedule 6-8 people together -Inaccurate recall
Direct Observation	To gather accurate information about how a program actually operates, particularly about processes	- Covers events in real-time - Can adapt to events as they occur - Covers context of events - Obtain insight into personal behavior and motives	-Can be difficult to interpret observations -Time consuming/ costly -Can influence behaviors of program participants

Data Collection Methods (con't)

Method	Overall Purpose	Advantages	Challenges
Document Review	To obtain impression of how program operates without interrupting the program; is from review of applications, finances, memos, minutes, etc.	-Get comprehensive and historical information -Doesn't interrupt program or client's routine in program - Information already exists - Few biases about information - Broad coverage over time.	- Often takes much time - Info may be incomplete - Need to be clear about what looking for - Data is restricted to what already exists - Can have reporting biases -Access might be blocked
Surveys, Checklists	To quickly and/or easily get lots of information from people in a non threatening way	- Can complete anonymously - Inexpensive to administer to many people - Easy to compare and analyze - Can get lots of data - Many sample questionnaires already exist	- Potentially inaccurate recall/feedback - Wording can bias client's responses - Are impersonal - May need sampling expert - Doesn't get full story - May need an ICR
Monitoring Data	Assess degree to which environmental impacts are occurring	- Yields good accountability and impact data	-Takes time to see results -Often beyond agency control -Variability in application of process -Different monitoring standards for acceptable levels of performance

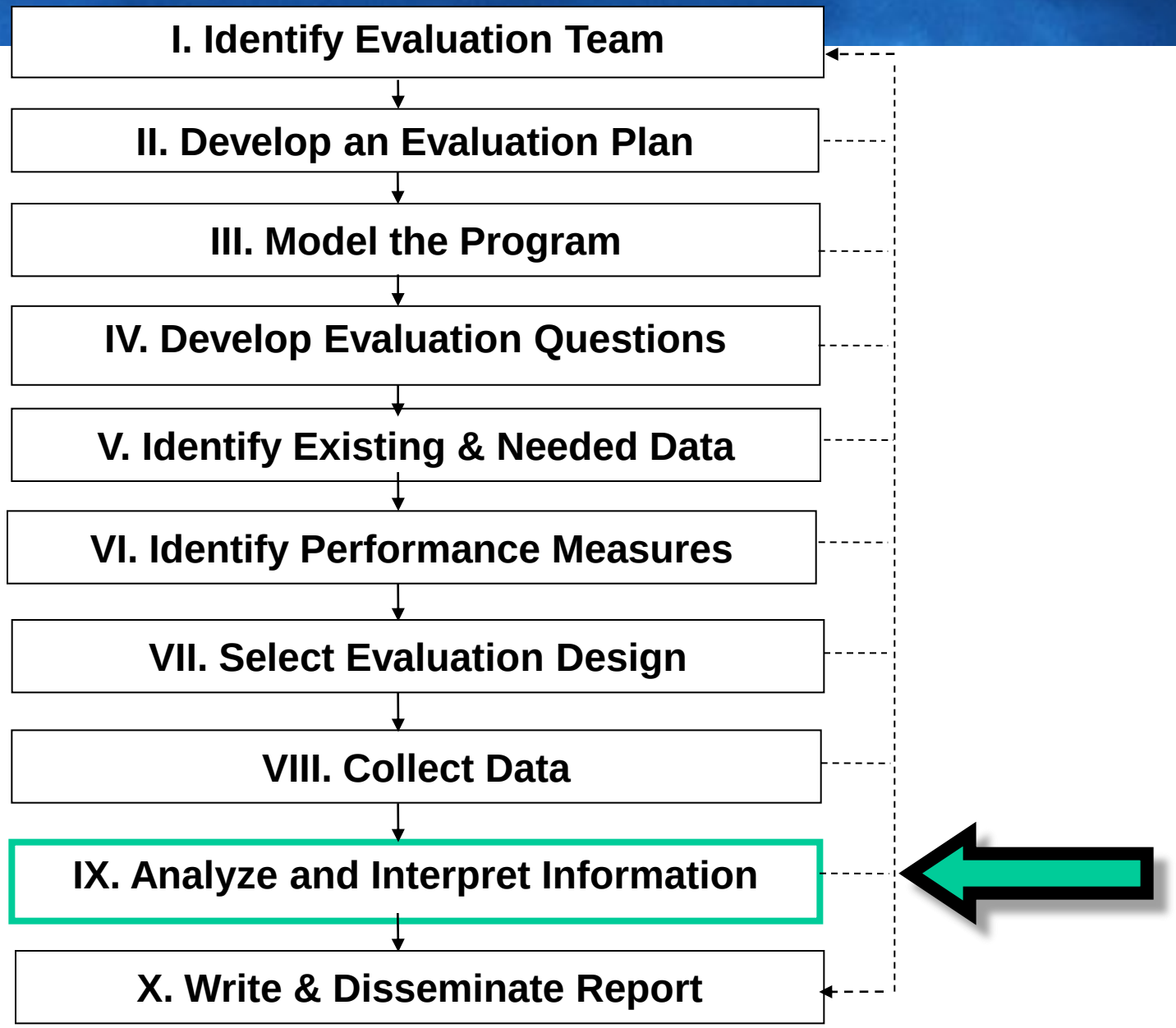
Tribal GAP Data Collection Methods

- Quantitative data for a sample of 111 tribes, drawn from four databases:
  - EPA GAP Accountability Tracking System (GAP database)
  - EPA Grants Information and Control System
  - EPA Strategic Goals Reporting System
  - US Census Audit Database
- File reviews conducted for tribes not included in GAP database to ensure representative data from all EPA regions
- GAP database and file review data represents GAP grantee activity from October 2000 – September 2004
- Qualitative data drawn from:
  - Interviews with GAP project officers in eight regions
  - Panel discussions with tribal representatives at three regional tribal meetings

Other Considerations for Data Collection

- **Field test** data collection instrument & data entry process.
  - Test your draft interview questions/ survey questionnaire on 2-4 people who are similar to the people from whom you'll be collecting data. Find out:
    - Did they understand the terms being used?
    - Did they interpret the terms as you intended?
    - Did the questions/ response scales make sense to them? Were they relevant?
- Determine if an ICR is needed
  - Under the Paperwork Reduction Act, federal agencies must have an OMB-approved Information Collection Request (ICR) to ask for identical information from 10 or more people who are not federal employees.

Steps to Completing an Evaluation



Data Analysis

In order to be meaningful, the raw data must be computed. Analysis involves comparison of the data to enhance interpretability.

Compare:

- Trends over time
- Actual performance against targets
- Variation across units (internal benchmarking)
- Against benchmarks (external benchmarking)
- With other breakouts

(From Performance Measurement for Government & Nonprofit Organizations - Dr. Theodore Poister- 7-23-24, 2004)

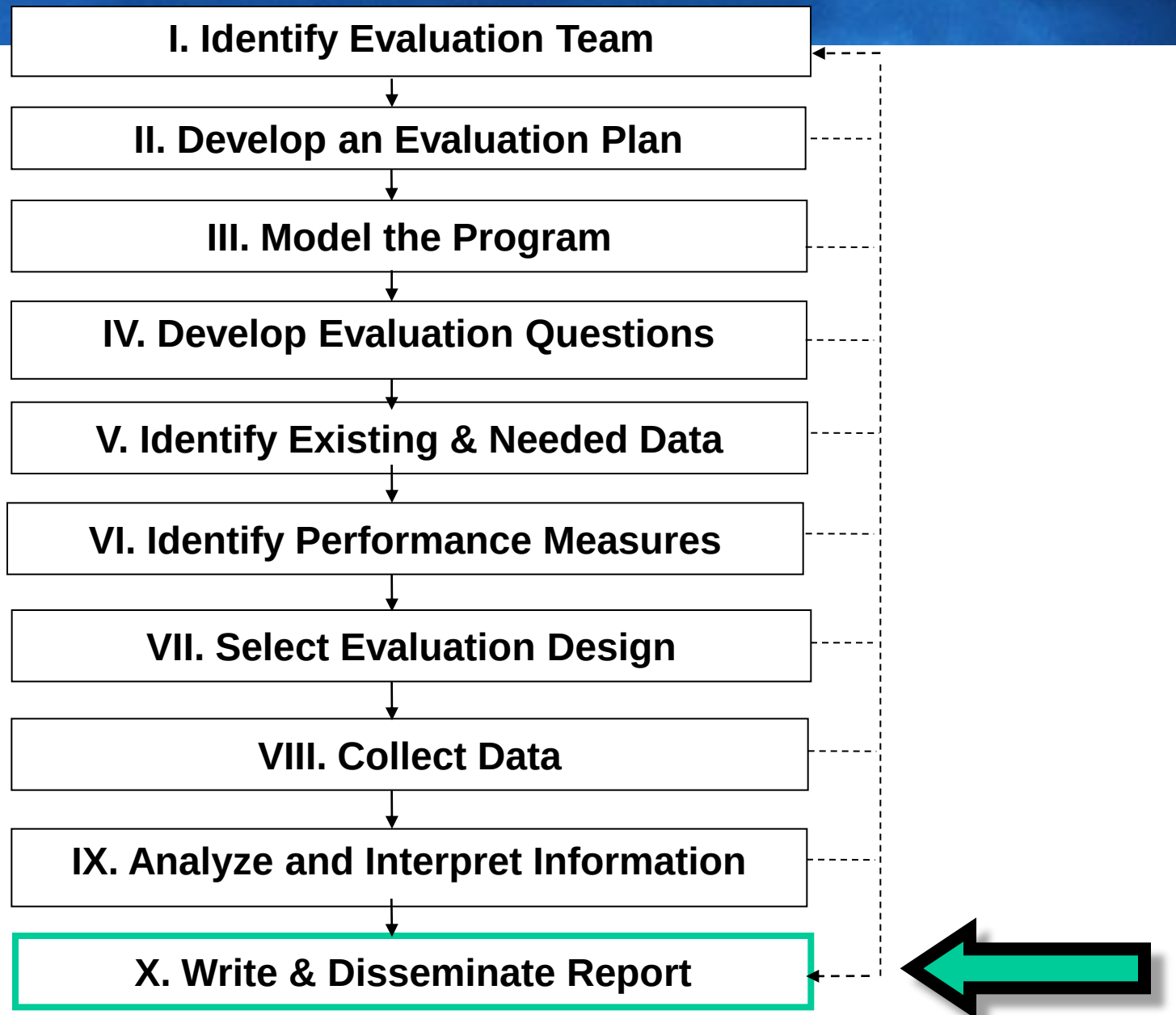
Analyzing Qualitative Data

- Content analysis
- Identify common, recurring patterns and themes, ideas, words or phrases the data
- Establish categories
- Code Data
- Analyze Data
  Bring order to the data
- Identify factors that will explain deviation
- Determine patterns/themes that will corroborate findings
- Interpret the Data
  Make sense of the data, find meaning and significance

Tips on Data Analysis

- Save sufficient time to thoroughly analyze the data.
- Analyze the data throughout the data collection process to catch problems as they arise.
- Analyze data in a manner that meets your audience's needs
- Tailor the analytical procedures to the type of data collected.
- Have more than one person review the analysis.

Steps to Completing an Evaluation



Writing the Report: Things to Consider

- Purpose
- Audience
- Timing
- Content
- Format
- Cost
- Communicating Negative Findings

Types of Reports

- Formal
  - Fact Sheet
  - Interim reports
  - Written report
  - PowerPoint briefing
- Informal
  - Internal memoranda
  - Verbal presentation
  - Personal reports

Components of the Report

- Executive Summary
- Background
- Methodology
- Findings
- Conclusions*
- Recommendations*
- Lessons Learned*
- Appendixes*

**** Optional***

Tips for Writing the Report

- Make sure the report is involved and active
- The report should be action oriented – focused on findings and recommendations, less on study design and analyses (these details can be in the appendices)
- The report should inform the audience of potential misuses
- Discuss methodological weaknesses & limitations
- Use case studies & interview/surveys to elaborate & provide a deeper understanding

Appendix A: Supplemental Information

Appendices

Appendix A: Supplemental Information

Appendix B: Managing the Evaluation

Appendix C: Working w/ A Contractor

Appendix D: Evaluation Resources

Types of Evaluation across the Life of a Program

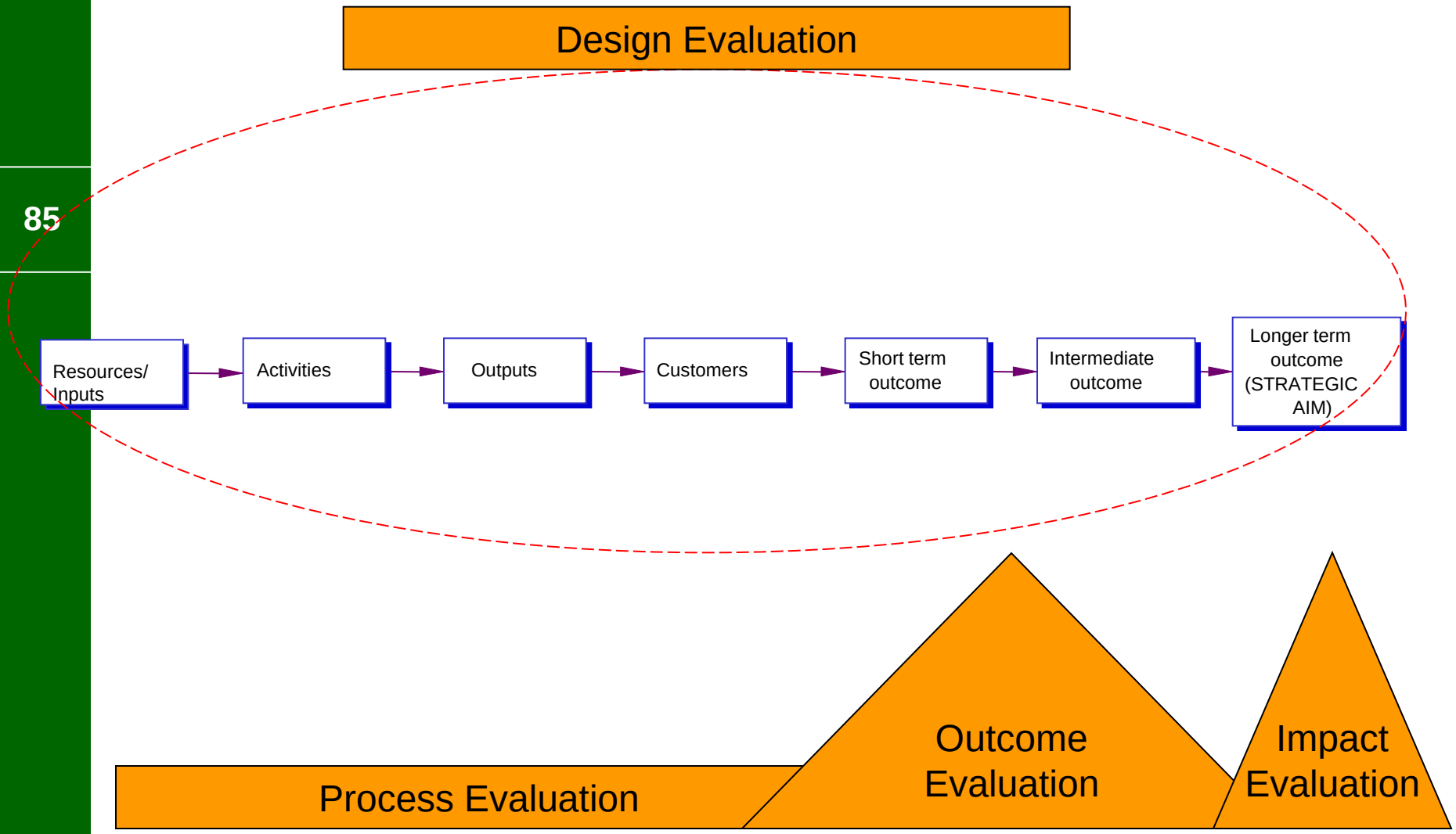
Evaluation Type	Program Life Cycle	Purpose	Typical Questions
Design/ Developmental Evaluation	**Developmental** (During initial stages of program development & design)	Instrumental in designing a new program and examining if it is conceptually sound	Where are we now? Where do we want to be? What are appropriate goals and objectives? Is the design of the program well formulated, feasible, and likely to achieve the intended goals?
Process/ Implementation Evaluation	I**mplementation** (During existing program implementation)	Focused on assessing process of program implementation, operations, service delivery	To what degree is the program strategy being implemented as intended? Are our processes and systems appropriate given our objectives?
Outcome Evaluation	**Mature Program or Program Completion** (Retrospective on mature program)	Examines program's short, intermediate & long-term outcomes	Were outcomes achieved? How can we improve our level of effectiveness? Were there unanticipated results?
Impact Evaluation		Focused on questions of program causality. Attempts to link the outcomes of the program to the changed or improved condition.	Did the program cause the desired outcome/impact?

Assessing Whether to Evaluate Your Program (Evaluability Assessment)

1. Is the program significant enough to merit evaluation?
    - Consider: program size, # of people served, transferability of pilot, undergoing PART
2. Is there sufficient consensus among stakeholders on program's goals and objectives?
3. Are staff & managers willing to make decisions about or change the program based on evaluation results?
4. Are there sufficient resources (time, money) to conduct an evaluation?
5. Is relevant information on program performance available or can it be obtained?
6. Is an evaluation likely to provide dependable information?
7. Is there a legal requirement to evaluate?

(Adapted from Worthen et al. 1997.)

Evaluation and the Logic Model



Adapted from Evaluation Dialogue Between OMB and Federal Evaluation Leaders: Digging a Bit Deeper into Evaluation Science, April 2005

Evaluation Questions Across the Performance Spectrum (Logic Model)

Program Elements:	Resources	Activities/ Outputs	Target Customer	Short-term Outcome	Intermediate Outcome	Long-term Outcome
Evaluation Questions:	▪ Which resources are most effective for achieving program objectives? Why?	▪ Are we executing the program as intended? ▪ Are we producing products & services at anticipated levels? ▪ Why or why not?	▪ Are we reaching the target audience? ▪ How can we improve customer satisfaction? ▪ What aspects of the program lead to greatest customer satisfaction?	▪ Were there changes in customers' attitude, knowledge, or skills? ▪ What role did the program have in observed changes? ▪ Which program services were most effective in generating change?	▪ How did customers' behaviors change as a result of the program? ▪ What aspects of the program were most effective in generating these behavioral changes?	▪ Did the program achieve its objectives? ▪ What changes in environment were achieved? ▪ Did the program create unintended outcomes? Were these beneficial, neutral, or detrimental to program objectives?
External Factors:	How did external factors influence my program's success? How could the program have mitigated for these effects?					

Some Definitions

- **Independent variable**: Variable assumed to be an antecedent condition to an observed behavior.
- **Dependent variable**: Variable not under the evaluator's control -- the data observed and measured in an evaluation in response to the independent variable.

➢ *In a cause-and-effect relationship, independent variable is the cause. Dependent variable is outcome or effect.*

- **Experimental group**: Group of subjects given the treatment / intervention / service / program.
- **Control group**: Group of subjects left unexposed to some procedure and then compared with treated subjects.

Adapted from Robert J. Hall, Associate Professor of Educational Psychology, Texas A&M University

Choosing an Evaluation Design

- Consider the type of evaluation being proposed:
  - **Design—**likely use non-experimental design
  - **Process—**likely use non-experimental design
  - **Outcome—**attempt to use quasi or true experimental designs
- Particularly for outcome evaluation, consider whether it would be feasible to:
  - Conduct a **pretest** (to obtain a baseline)
  - Compare with a **control** group
  - **Randomly** assign subjects to control and treatment groups

Types of Evaluation Designs

I. Non-Experimental	II. Quasi-Experimental	III. True Experimental
Pretest/posttest: Sometimes Control Group: No Random Assignment: No	Pretest/posttest: Yes Control Group: Usually Random Assignment: No	Pretest/posttest: Yes Control Group: Yes Random Assignment: Yes
One-Shot Design (quantitative) ***or* Case Study (qualitative)** X → O	**Non-Equivalent Control Group (with pretest/posttest)** Group A: O → X → O Group B: O O	**Pre/Posttest Control Group** Group A: R → O → X → O Group B: R → O O
One Group Pretest/Posttest O → X → O	**Single Group Interrupted Time Series** O-O-O-O-O → X → O-O-O-O-O	**Posttest Only Control Group** Group A: R → X → O Group B: R O
Posttest-Only Control Group Group A: X → O Group B: → O	**Control Group Interrupted Time Series** A: O-O-O-O-O → X → O-O-O-O-O B: O-O-O-O-O O-O-O-O-O	**Solomon Four-Group** Group A: R → O → X → O Group B: R → O O Group C: R X → O Group D: R O
Alternative Treatments Group A: Xa → O Group B: Xb → O	**Key:** **X = Treatment / program applied** **O = Observation / measurement** **R = Random assignment (each subject has equal chance of being selected into control or treatment group)**	

I. Non-Experimental Designs

Examine changes in outcomes of program participants but does not compare to groups not exposed to program. Weak in controlling for internal and external validity.

- One-Shot Design
  - Measurement takes place one time only, following the intervention or program (e.g. a post-workshop survey)
- Case Study Design
  - Involves in-depth data collection and analysis of a particular individual, group, or program.
- One Group Pretest-Posttest
  - Measurement takes place both before & after intervention
- Posttest Only Control Group
  - Data collected on two separate groups (treatment group & control group) only after the intervention or program.
- Alternative Treatments
  - Data collected on two groups (given different interventions) only after the interventions.

II. Quasi-Experimental Designs

Compare outcomes from program participants to outcomes for comparison/control groups that do not receive program services. Groups are not randomly assigned. Stronger than non-experimental designs in controlling for internal and external validity.

- Non-Equivalent Control Group (pre & post-test)
  - Before and after measurements taken from both control and treatment groups.
- Single Group Interrupted Time Series
  - Measurement gathered on a single group both before and after implementation of the intervention or program
- Control Group Interrupted Time Series
  - Same as single group design, except that measurements are also gathered from a control group that did not experience the intervention or program

III. True Experimental Designs

Compares outcomes from program participants to outcomes for comparison groups that do not receive program services. Randomly assigns participants into control & treatment groups. Strongest design in controlling for internal and external validity.

- Pre/Posttest Control Group
  - Data are collected on two separate groups (treatment group & control group) both before and after an intervention or program.
- Posttest-Only Control Group
  - Data collected on two separate groups (treatment group & control group) only after the intervention or program.
- Solomon Four-Group
  - Data collected on four groups to observe effects of intervention and pretesting. To observe effect of intervention, two groups receive pretest/posttest, but one (control) does not receive the intervention. To observe effect of pretesting, two groups receive only the posttest, but one (control) does not receive intervention.

What Difference Does the Design Make?

A good design increases internal & external validity.

- **Internal Validity**: Extent to which the evaluation proves that the program, as opposed to other factors, created the difference we observe in the subjects.
- **External Validity**: Extent to which evaluation results can be generalized to other people and settings.

However, certain factors limit our ability to confidently claim:

- *whether our program made the difference (internal validity)*
- *whether our evaluation findings can be applied to other people and settings (external validity)*

Factors that Reduce Internal Validity

1. **History**: External events occurring during the course of the evaluation can affect results.
   - e.g., Hurricane Katrina would create history effects in an evaluation of water quality changes in New Orleans in 2005.
2. **Maturation**: Results change over the course of the evaluation simply because subjects have been in the experiment longer.
   - e.g., Participants become tired after an all-day Evaluation training and their posttest scores are affected.
3. **Pre-testing**: The actual process of being measured before a treatment (program) can affect how subjects respond to later measurements.
   - e.g., Participants in an Evaluation training take a pre-test and then watch for key concepts during the training session. Their posttest scores are higher than what they otherwise would have been.

Factors that Reduce Internal Validity (con't)

5. **Instruments**: Changes in the measuring instrument or observers causes changes in before & after measurements.
   - E.g., A survey is used to rate a facility's knowledge of P2 techniques. The treatment (technical assistance) is applied, followed by the same survey, but new questions were added.
6. **Statistical Regression**: Groups chosen due to extreme scores revert to the mean with repeated measurements.
   - E.g., Poor performing facilities are given technical assistance. Average posttest scores will be higher even without the TA.
7. **Selection**: People are pre-selected into control and treatment (program) groups that are not equivalent.
   - E.g., A group of people given tips on energy conservation is compared with a group that is not. However, there was no way of knowing if the groups were equivalent to begin with.
8. **Mortality**: Because of unique characteristics, loss of subjects from treatment (program) or control groups can affect results.
   - E.g., An evaluation of a pilot program starts with 10 facilities but ends with 5. Remaining facilities may be different from those that left.

Factors that Reduce External Validity

1. **Pre-testing**: The actual process of being measured before a treatment (program) can affect how subjects respond to later measurements.
   - E.g., Participants in Evaluation training take a pre-test and then watch for key concepts during the training session. Posttest scores are higher than what they otherwise would have been.
2. **Differential Selection**: How subjects are selected into the treatment group influences how generalizable their findings are to larger populations.
   - E.g., People who are browsing a recycling website and take an online survey will likely be different than the general population.

Factors that Reduce External Validity (con't)

3. **Experimental Procedures ("Hawthorne Effect")**: The process of undergoing any experimental procedures can make subjects respond.
   - E.g., Light manipulation both up and down increased worker productivity in the study of the Hawthorne plant.
4. **Multiple Treatment Interference**: Observed outcomes from subjects given more than one treatment cannot be isolated to a particular treatment.
   - E.g., Facilities are given technical assistance via workshops and a handbook. Outcomes cannot be isolated to any one treatment.

Evaluation Design Features that Increase Internal and External Validity

1. **Pre-testing**: Taking a measurement before the program/treatment is applied can set a baseline and help determine the effect of the treatment.
   - *However, pre-testing can be another threat to validity.*
2. **Control Group:** Using a similar group that is not exposed to the treatment/ program can reduce History, Maturation, and Instrumentation threats.
3. **Randomization**: Randomly selecting subjects can help reduce Statistical Regression and Differential Selection threats.
4. **Additional Groups**: Adding additional groups who do not receive the Pretest or Experimental Procedures can control for those particular threats.

Information Collection Requests (ICRs)

- Under the Paperwork Reduction Act, federal agencies must have an OMB-approved Information Collection Request (ICR) to ask for identical information from 10 or more people who are not federal employees.
- An ICR:
  - Describes information to be collected.
  - Gives reason the information is needed.
  - Estimates time and cost for public to answer the request.
- The ICR process takes at least 9 months.
- EPA's ICR Center: http://intranet.epa.gov/icrintra/index.html

Tips on Planning for Data Collection

- Weigh pros and cons of different methods. (Slides 49-50)
- Use a method that answers multiple evaluation questions.
- Consider how data will be analyzed **BEFORE** collecting it!
  - If considering a survey, consult with a statistician to ensure that the questions are designed appropriately.
  - Assume hypothetical data have been collected and note whether they answer the evaluation questions well.
- Constantly refer back to your evaluation questions. Are you answering the questions with these methods?
- Consider software needs (for compiling & analyzing data).
- **Field test** data collection instrument & data entry process.

Field Testing

- Test your draft interview questions/ survey questionnaire on 2-4 people who are similar to the people from whom you'll be collecting data. Find out:
  - Did they understand the terms being used?
  - Did they interpret the terms as you intended?
  - Did the questions/ response scales make sense to them? Were they relevant?
- Field test sampling and data entry processes
  - Especially if different people will be collecting and/or entering the data.
- **Revise** your data collection instrument or process based on the field test results.

Appendix B: Managing the Evaluation

- Before You Start the Evaluation:
  - Secure management buy-in prior to undertaking an evaluation.
  - Don’t rush the set up/scoping portion of the evaluation.
  - Be realistic about time and resource (staff, $$) constraints.
  - Narrow the scope of the evaluation if needed.
  - Involve management early on in framing the scope of the evaluation.
- During the Evaluation:
  - Ensure consistent and extensive involvement by key staff
  - Be aware of cultural and political sensitivities.
  - Communicate the results to all stakeholders routinely to avoid surprises.

Appendix B: Managing the Evaluation

- Monitor the evaluation project's tasks and personnel
- Ensure the evaluation stays on schedule or negotiate schedule changes
- Monitor the evaluation's budget
- Ensure the evaluation stays on budget or negotiate changes
- Keep the client and stakeholders informed of progress and problems

Appendix C: Working With the Contractor

- Select contractors that have evaluation experience.
- Choose a contract vehicle that allows uninterrupted service and access to contractors with evaluation expertise.
- The contractor will develop all products specified in the work assignment/work plan.
- EPA staff should work with the contractor to facilitate data collection from internal and external evaluation stakeholders.
- Direct involvement by the program office leads to a better report that is more likely to meet the needs of the program and whose recommendations are more likely to be implemented.
- To ensure objectivity, the contractor should take responsibility for evaluation conclusions and drafting the final report.

Appendix D: Evaluation Resources Websites

- **W.K. Kellogg Foundation's Evaluation Toolkit:** Contains resources on developing evaluation questions, plans, budgeting for evaluation, managing a contractor, and checklists. Includes the Evaluation Handbook and Logic Model Development Guide. http://www.wkkf.org/default.aspx?tabid=75&CID=281&NID=61&LanguageID=0
- **National Science Foundation:** User-friendly handbook for evaluations integrating quantitative and qualitative methods. http://www.ehr.nsf.gov/EHR/REC/pubs/NSF97-153/start.htm
- **US General Accounting Office:** GAO policy and guidance materials on evaluations, designing evaluations, case study evaluation, and prospective evaluation methods. http://www.gao.gov/policy/guidance.htm
- **The Evaluation Center at Western Michigan University:** Excellent resource for evaluation checklists, instructional materials, publications, and reports. http://www.wmich.edu/evalctr/

Evaluation Resources: Websites (con't)

- **American Evaluation Association:** Professional society for evaluators with links to evaluation websites. **http://eval.org**
- **Online Evaluation Resource Library:** Contains evaluation instruments, plans, reports, and instructional materials on project evaluation design and methods of collecting data. **http://oerl.sri.com/**
- **The Evaluator's Institute:** Offers short-term professional development courses for practitioners. **http://www.evaluatorsinstitute.com**
- **Collaborative & Empowerment Evaluation** website: **http://www.stanford.edu/~davidf/empowermentevaluation.html**
- **SurveyMonkey:** Free online survey package. **www. surveymonkey.com**
- **Center for Disease Control** Evaluation Resources: **http://www.cdc.gov/healthyyouth/evaluation/resources.htm**

Evaluation Resources: Websites (con't)

- **Conservation Measures Partnership:** A partnership of conservation groups whose goal is to improve the design, mgmt, and measurement of conservation action. They have expanded the adaptive mgmt cycle in their *Open Standards for the Practice of Conservation* and have developed the Miradi software for use in project design and implementation, including logic model development. **www.conservationmeasures.org**; **www.miradi.org**
- **MEERA (My Environmental Education Evaluation Resource Assistant)** MEERA's goal is to support the evaluation efforts of environmental educators. MEERA seeks to meet this goal by facilitating access to relevant information and resources through a single, web-based location: **http://meera.snre.umich.edu/ http://meera.snre.umich.edu/** . MEERA features step-by-step guidance for planning & implementing evaluations, reviews of how to guides specific to EE & other evaluation contexts, reviews of evaluation websites & resources, a database of EE evaluation reports.

Evaluation Resources: Publications

- *Handbook of Practical Program Evaluation.* Woley, J., Hatry P., and K. Newcomer. 1994. San Francisco: Jossey-Bass Publishers.
- *Evaluation: A Systematic Approach*, Rossi, P., Freeman, H., and M. Lipsey. 1999. Thousand Oaks, CA: SAGE Publications.
- *Evaluation in Organizations: A Systematic Approach to Enhancing Learning, Performance, and Change.* 2001. Russ-Eft, D. and H. Preskill. Cambridge, MA: Perseus Publishing.
- *Program Evaluation: Alternative Approaches and Practical Guidelines*. 2nd ed. Worthen, B., Sanders, J., and J. Fitzpatrick. 1997. New York: Addison Wesley Longman.

Evaluation Resources: Publications (con't)

- *The Manager's Guide to Program Evaluation: Planning, Contracting, and Managing for Useful Results*. Mattessich, P. 2003. Saint Paul, MN: Wilder Publishing Center.
- *Real World Evaluation: Working Under Budget, Time, Data, and Political Constraints.* Bamberger, M., Rugh, J. and L. Mabry. 2006. Thousand Oaks, CA: Sage Publications.
- *Evaluability Assessment: A Practical Approach.* Smith, M. 1989. Norwell, Mass: Kluwer.
- *Utilization-Focused Evaluation: The New Century Text.* 3rd ed. Patton, M. 1997. Thousand Oaks, CA: Sage Publications.