

# The Evaluation Toolkit

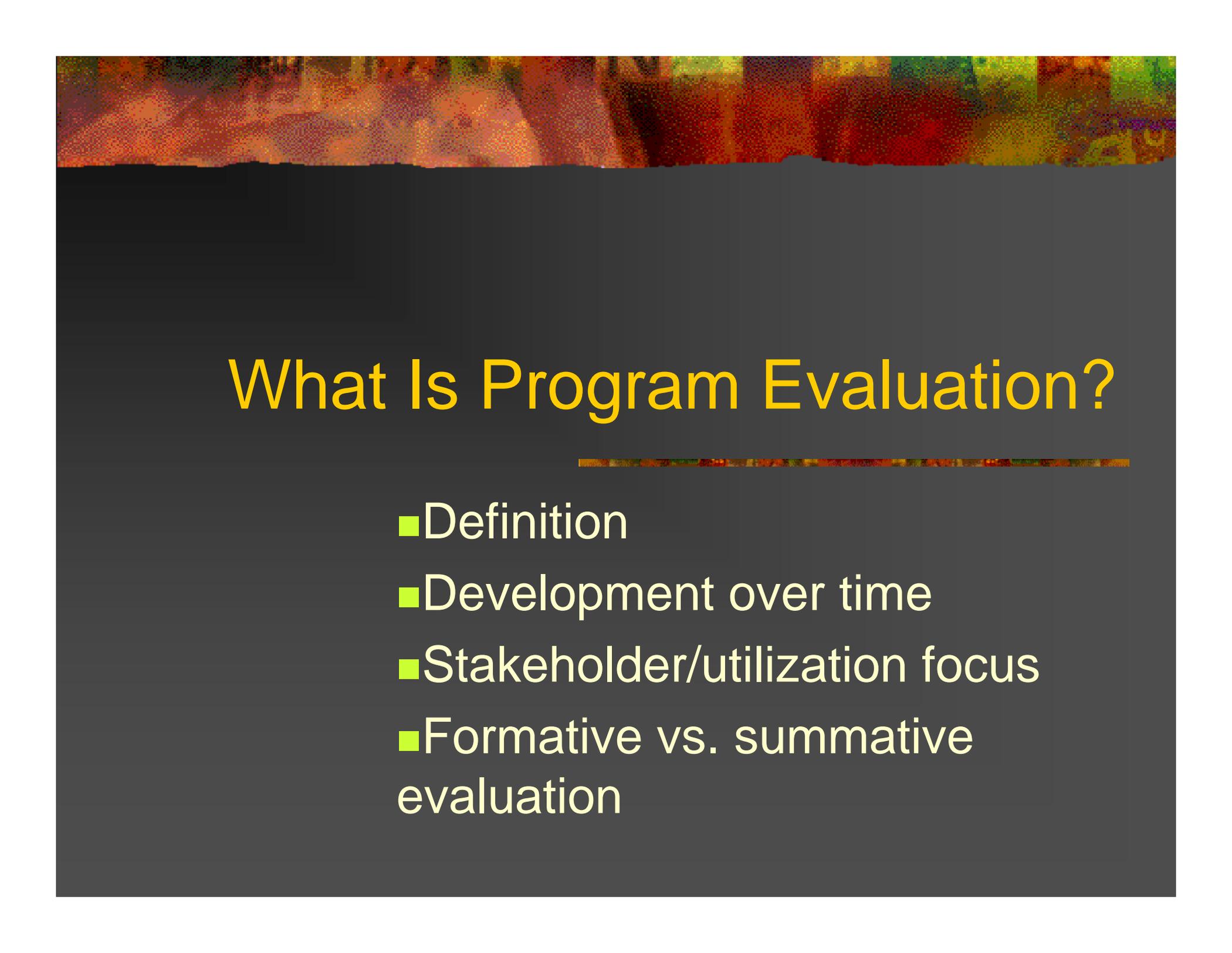
---

Basic Concepts and Practices

# Overview

---

- What is evaluation?
  - The Program Evaluation Standards
  - The CIPP Model
    - Context evaluation
    - Input evaluation
    - Process evaluation
    - Product/outcome evaluation
-



# What Is Program Evaluation?

---

- Definition
- Development over time
- Stakeholder/utilization focus
- Formative vs. summative evaluation

# What Is Program Evaluation?

## *Definition*

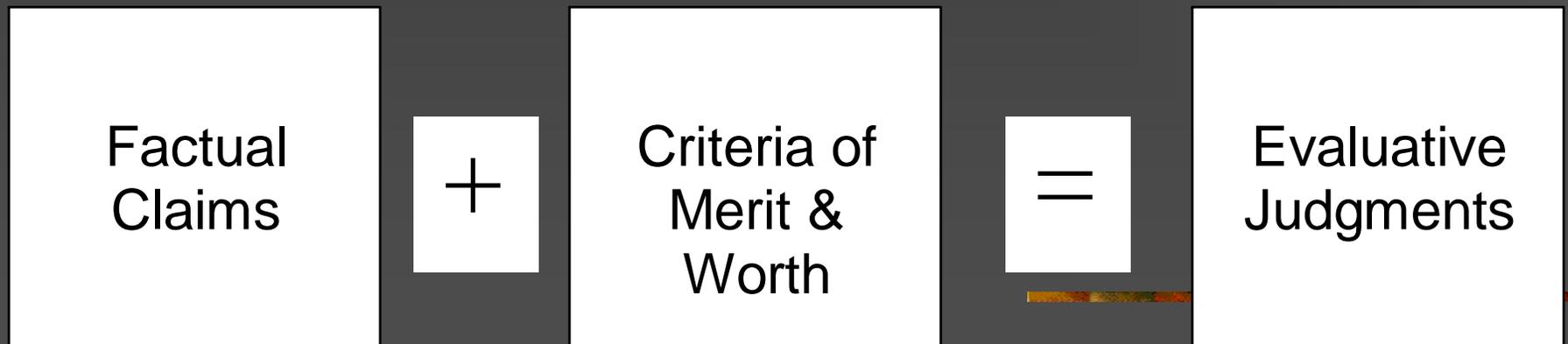
---

- The *systematic* assessment of the *merit* or *worth* of a program (Joint Committee, 1994)
    - *Systematic* – inferences are arrived at based on sound reasoning and use of evidence
    - *Merit* – the intrinsic value of a program
    - *Worth* – the value of a program in a given context
-

# What Is Program Evaluation?

## *Definition (2)*

- Evaluation involves determining merit & worth
- Thus, it is a synthesis of facts & values
- Evaluation cannot be value free
- Criteria of merit and worth must be justified



# What Is Program Evaluation?

*Evolution of viewpoints*

---

- Evaluation is a pervasive activity
- It is trans-disciplinary and multi-disciplinary
- The advent of program evaluation as a profession came about during the 1960s, with the increase in change programs

# What Is Program Evaluation?

## *Evolution of viewpoints (2)*

- Early evaluation was essentially *applied social science*
  - Employed traditional criteria of scientific merit (e.g., validity, reliability, rigor)
  - Took official program goals as evaluative criteria (“goal-based evaluation”)
  - Sought to keep stakeholders at arm’s length
  - Generally employed at the end of the policy/program development cycle
  - Focused on outputs and outcomes

# What Is Program Evaluation?

*Evolution of viewpoints (3)*

- By the mid-1970s there was widespread concern over lack of evaluation use
- Also, frustration with prevalence of null findings
- Responses
  - Stakeholder orientation
  - Formative vs. summative

# What Is Program Evaluation?

## *Stakeholder orientation*

- Engage stakeholders in evaluation design and execution (utilization focus)
  - Gain buy-in
  - Enhance use
  - Ease evaluation anxiety by developing consensus on evaluation use (and misuse)
  - Capitalize on stakeholders' informal know-how, in addition to scientifically validated information

# What Is Program Evaluation?

*Formative, as well as, summative use of evaluation*

- Summative evaluation
  - After completion of the program cycle
  - Generally for an external audience
  - “When the customer tastes the soup . . . .”
- Uses
  - Accountability to external stakeholders
  - “Recycling” decisions

# What Is Program Evaluation?

*Formative, as well as, summative use of evaluation (2)*

- Formative evaluation
  - In-process
  - Generally for an internal audience
  - “When the cook tastes the soup . . . .”
- Uses
  - Program planning
  - Mid-course corrections
  - Data points for internal discussion

*The most important purpose of evaluation is not to prove  
but to improve -D. Stufflebeam*

# What Is Program Evaluation?

*Reconciling the views*

---

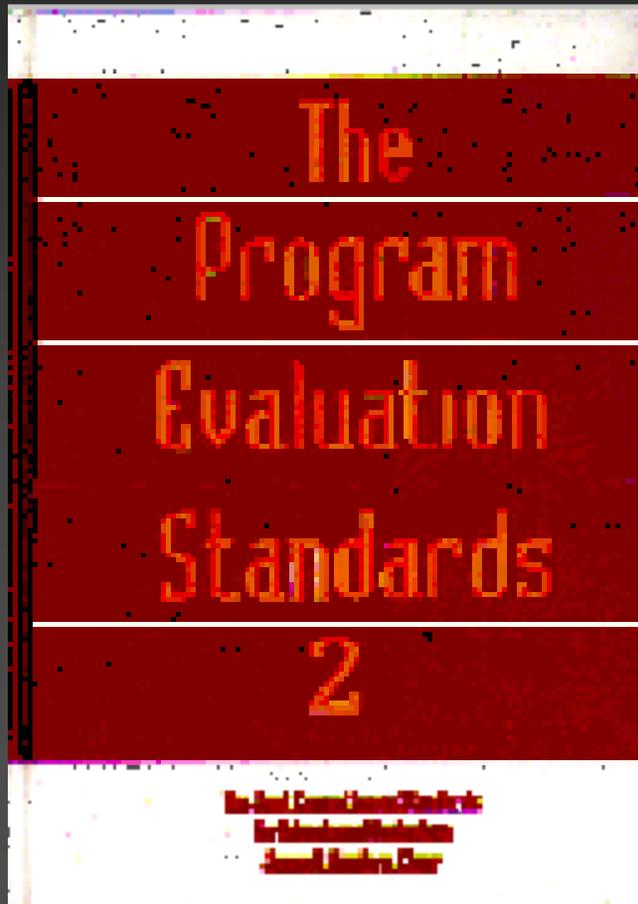
- There is a persistent debate over the appropriateness of summative vs. formative evaluation and appropriate role of stakeholders
  - This workshop focuses mainly on formative evaluation
  - Our purpose is not to discredit summative evaluation, but to round out the evaluation toolkit
-

# What Is Program Evaluation?

## *The Bottom Line*

- Evaluation is research plus “something more”
- Being a good evaluator requires solid research skills, but also
  - Facility with value questions
  - Ability to work with stakeholders
  - Strong communication skills
  - Good instincts about the use context
- See handout on research vs. evaluation

# The Program Evaluation Standards



# The Program Evaluation Standards

*Operationalizing the conception of evaluation*

- Originally developed for evaluating educational programs
- Developed by a joint committee with representatives of a variety of professional organizations
- More recently adopted by a wide range of evaluators
- Note the only set of evaluation standards around (cf. AEA's Guiding Principles). But both sets address similar concerns

# The Program Evaluation Standards

*Four core standards*

---

## ■ *Utility (U)*

- intended to ensure that an evaluation will serve the information needs of intended users

## *Feasibility (F)*

- intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal
-

# The Program Evaluation Standards

## *Four core standards*

---

- *Propriety (P)*

- 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 (A)*

- 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
-

# The Program Evaluation Standards

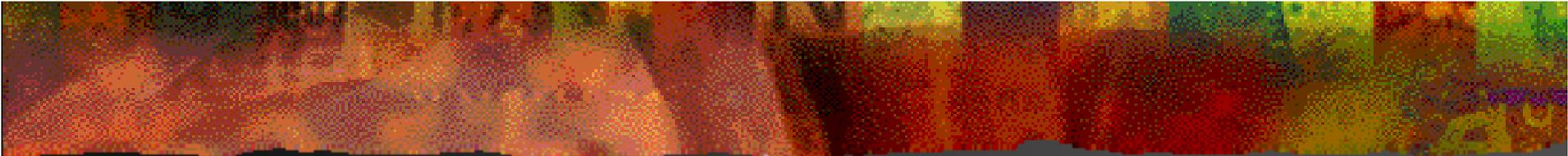
## *Discussion*

- What are some barriers to effective use of evaluation in your work?
- What are the barriers you have experienced to the effective use of your evaluation or research reports?

# The Program Evaluation Standards

## *Concluding thoughts*

- Difficulty of managing political volatility often increases with number of groups, salience of issue
- At some point it might be better to discontinue an evaluation than to allow it to fall prey to political machinations
- There is no substitute for detailed knowledge of stakeholders and issues



# The CIPP Model

---

- Context
- Input
- Process
- Product

# The CIPP Model

*Embedding evaluation in all stages of the program development cycle*

- A template of issues and questions to examine
- Compatible with the full range of research methods (qualitative and quantitative)
- Key elements
  - Context
  - Input
  - Process
  - Product

# The CIPP Model

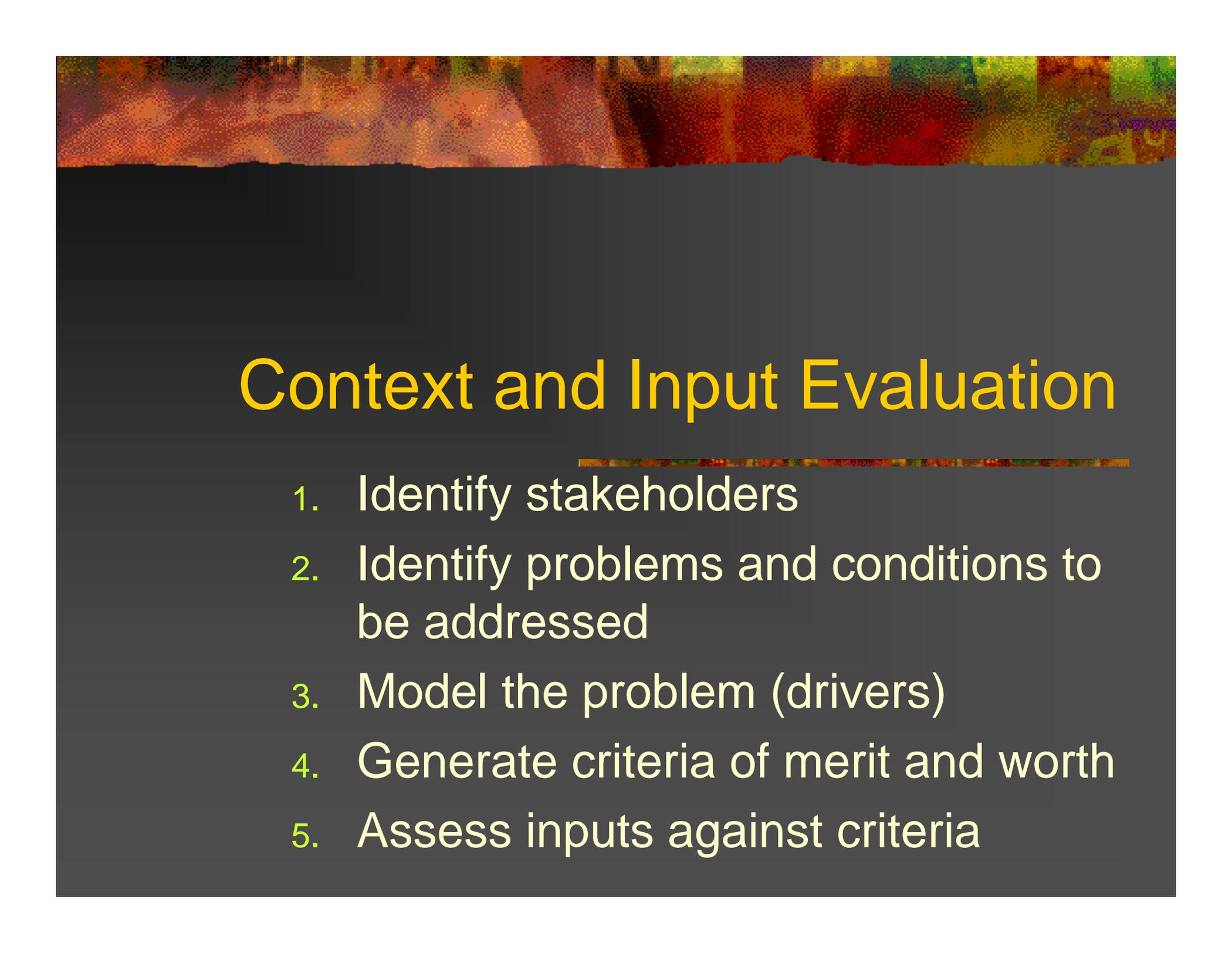
*Uses in formative and summative evaluation*

- Uses in formative evaluation
  - Planning, needs assessment (context evaluation)
  - Program design (input evaluation)
  - Improving implementation (process evaluation)
  - Assessing impacts (product evaluation)

# What Is Program Evaluation?

*Formative vs. summative evaluation*

- Uses in summative evaluation – improve utility (actionability) of findings about impact
  - Program goals are sometimes ill-considered (Context)
  - Might be a good program, but not well-matched to this particular context (Input)
  - Poor implementation (Process)
  - All of these factors can confound impact evaluations, limiting their ability to provide actionable knowledge!



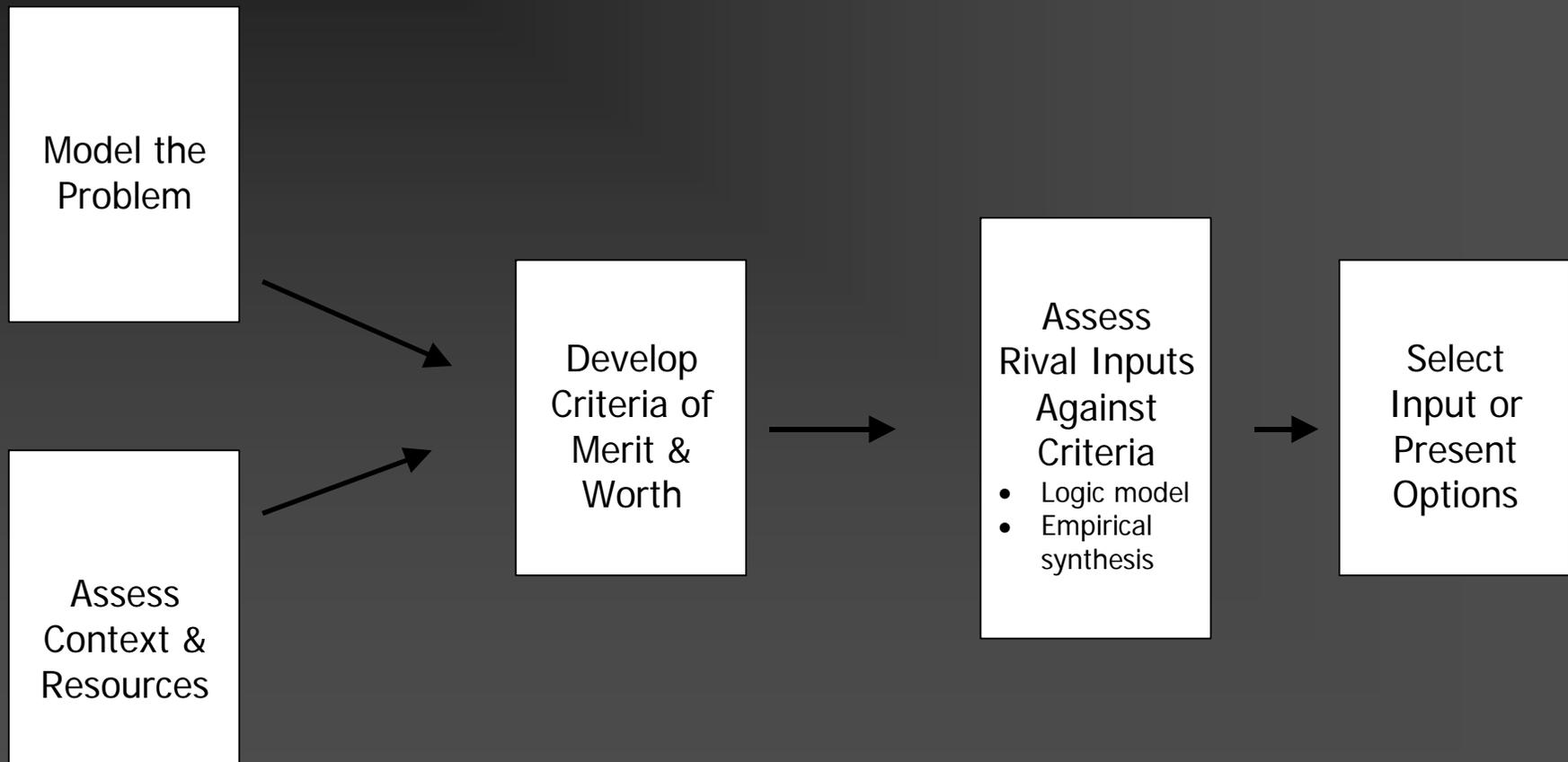
# Context and Input Evaluation

1. Identify stakeholders
2. Identify problems and conditions to be addressed
3. Model the problem (drivers)
4. Generate criteria of merit and worth
5. Assess inputs against criteria

# Background

- Organizations often find it surprisingly difficult to set priorities and select programs in a rational manner
  - Garbage can model and the problem of parallel streams
- Beware of blithely accepting stated program goals as evaluative criteria
- Context evaluation helps determine the appropriateness of program goals
- Input evaluation assesses the fit between appropriate goals and program designs

# Overview of Key Steps



# Identifying Stakeholders

---

- Clients vs. stakeholders
  - Strategies
    - Begin with the intended target population but be prepared to go beyond it
    - Try to anticipate those affected by spillover costs and benefits
    - Use group leaders to identify other stakeholders
-

# Identifying Stakeholders

---

- Common problems and issues
    - Clients sometimes have restrictive views of who counts as a “stakeholder”
    - Evaluators sometimes blithely assume that client is the only stakeholder, or accept target population as only stakeholder
-

# Identifying the Problem

- Identify general program domain
- *What is the behavior or condition in need of changing?*
  - Legal mandates
  - Needs assessment
  - Stakeholder consensus
  - Public agenda
- At this stage, avoid defining problem in terms of means (e.g., absence of money)

# Identifying the Problem

---

- *How bad is the problem?*
    - Are there likely to be changes in the the incidence and prevalence of the problem in the future?
    - How much uncertainty is there about current and future incidence/ prevalence?
    - Try to obtain worst- and best-case estimates
    - Is there heterogeneity in the target population?
-

# Modeling the Problem

## *Identifying drivers*

---

- *What are the key causes/drivers of the problem?*
  - Are the causal processes (drivers) different for different sub-populations?
  - Does the strength of the causal processes (drivers) differ across subpopulations?
-

# Assessing Context & Resources

## *Implementers & Target Population*

---

- “Coproductive” resources
    - Fiscal resources
    - Skills and capacities (coproduction)
    - Organizational climate and culture
  - Anticipate implementation issues (cf. process evaluation)
-

# Generic Criteria of Merit & Worth For Selection in Inputs

---

- Is input (program) design congruent with problem and context?
  - Does input seek to influence key leverage points?
  - Is the input likely to target the “right” people?
  - Is the input powerful enough given the severity of the problem?
  - Are the benefits likely to come in a timely fashion?

# Generic Criteria of Merit & Worth (2)

---

- Is input design congruent with problem and context? (cont'd)
    - Congruence between input and implementer attributes (wills, skills, and bills)
    - Congruence between input and target population attributes (wills, skills, and bills)
-

# Context and Input Evaluation

## *Feedback activities*

---

- Periodic reports on key indicators of relevant behaviors, conditions, resources, and constraints
  - Written reports
  - Briefings
  - Advocacy teams
  - Often, data collection activities themselves provide feedback
-

# Examples of Context Evaluation

---

- Formative example
- Summative example



# Exercise Using Context Evaluation

---

Mike Coplen



# Process Evaluation

---

1. Ascertain plan of action  
(evaluative criteria)
2. Identify key implementers
3. Data collection and analysis
4. Diagnose reasons for problems
5. Feedback

# Ascertain Plan of Action

---

- Key question involves the extent to which the program is being implemented as planned
    - Criterion of merit is plan of action
  - However, consider whether implementation plan is reasonable given goals and context
-

# Ascertain Plan of Action

---

## ■ Sources

- Program documents
  - Interview program designers
  - Seek norms of roll-out and scale-up (including site-specific targets)
  - Review literature on implementation of similar programs (if available)
-

# Identify Key Implementers

---

- When ascertaining plan of action note who is supposed to do what
- Beware that actual roles often deviate from planned roles

# Data Collection and Analysis

---

- Data collection
  - Interviews and focus groups
  - Observations
  - Document analysis
  - Surveys

# Data Collection and Analysis

## ■ Analysis

- Comparisons to norms
- Comparisons to site-specific goals
- Cross-sectional variation – are some doing a better job than others
- Temporal variation – are things getting better/worse over time
- Pay attention to whether variation in implementation is explicable in terms of task type, resource requirements, etc

# Diagnose Reasons For Problems

---

- Formative uses of process evaluation should both document problems and provide improvement-oriented feedback
- Requires diagnosis of barriers and supports

# Diagnose Reasons For Problems

- No substitute for detailed knowledge of program and context
- However, it is also useful to keep several typical barriers in mind
  - Wills
  - Skills
  - Bills
  - Systems
- Use findings of context evaluation!

# Diagnose Reasons For Problems

---

## ■ Wills

- Do implementers have appropriate incentives, positive attitudes, clear understanding of roles and responsibilities?
- Strength and clarity of program “signal”

## ■ Skills

- Do implementers know how to do what is required of them?
-

# Diagnose Reasons For Problems

---

## ■ Bills

- Do implementers have the resources to do what is required of them (money, time, human capital)

## ■ Systems

- Are organizational processes and structured congruent with program demands (distribution of authority, chain of command, social capital)
-

# Feedback

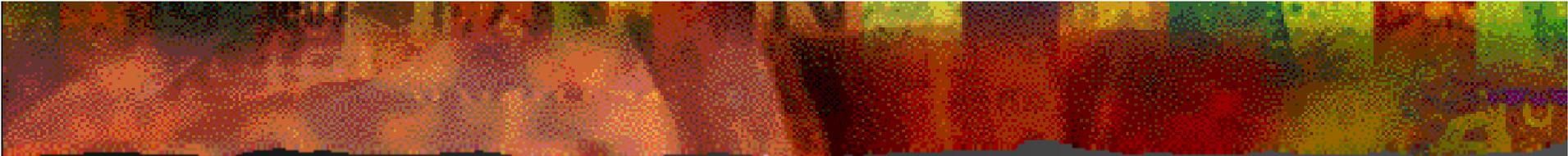
---

- Periodic reports and briefings to program managers
  - Feedback workshops with implementers
  - Consider pre-formative “concept-testing” exercise
    - Are co-productive assumptions of program congruent with resources and barriers?
    - Cf. Context and input evaluation
-

# Feedback

---

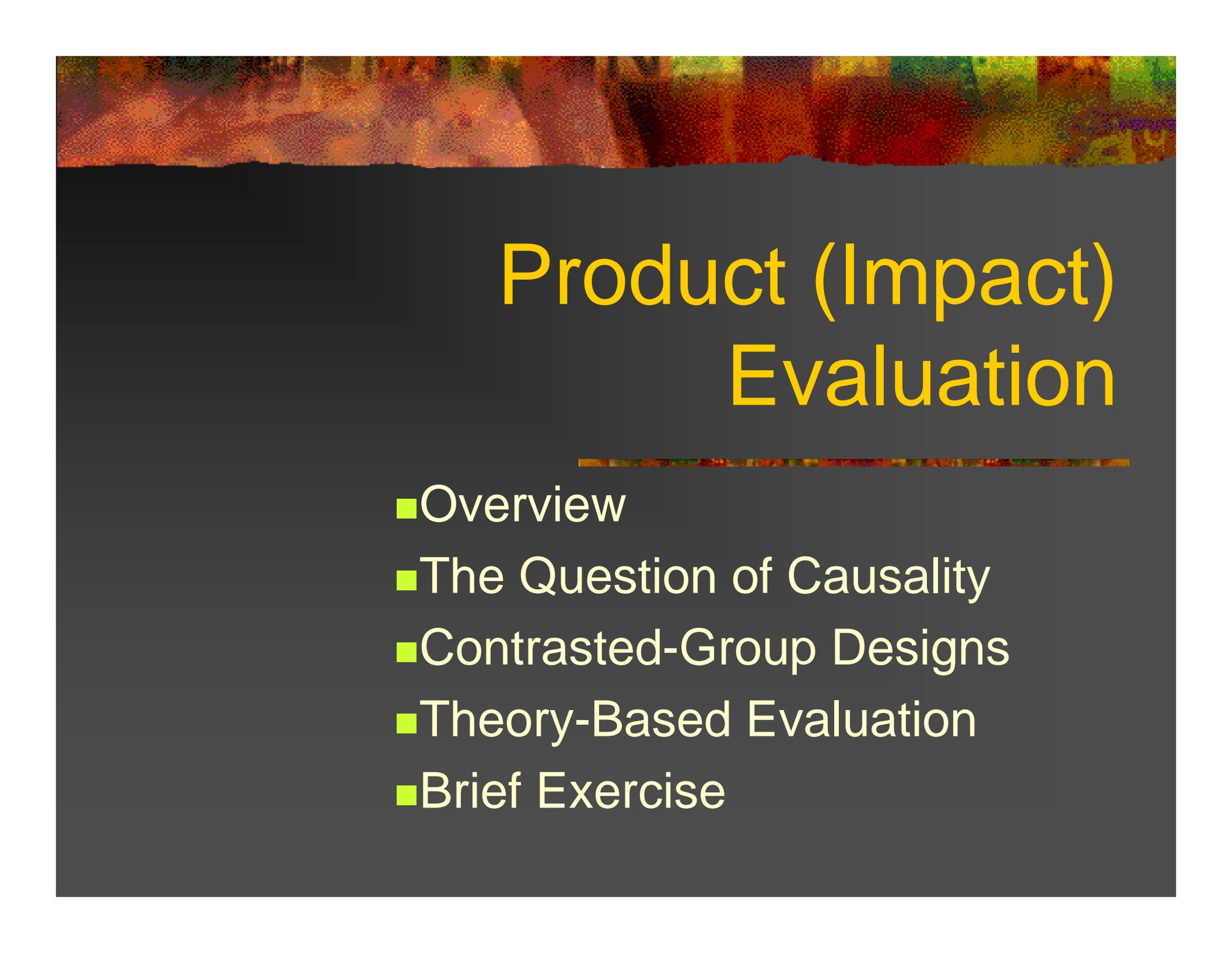
- Formative and summative example of process evaluation – RAND's evaluation of NAS



# Exercise Using Process Evaluation

---

Joyce Ranney



# Product (Impact) Evaluation

---

- Overview
- The Question of Causality
- Contrasted-Group Designs
- Theory-Based Evaluation
- Brief Exercise

# Overview

---

- Production evaluation is synonymous with assessment of
    - Impact
    - Effectiveness
  - Types of questions
    - Did performance meet standards?
    - What value was added by the program?
-

# Overview

---

- Product evaluation is usually a summative activity
- However, it can also be used formatively to provide ongoing feedback on performance indicators

# The Question of Causality

---

- Question of value-added raises the issue of causality
- Strongest product evaluations explicitly compare actual performance with evidence on the counter-factual
  - What would have happened in the absence of the program?

# Contrasted Group Designs

---

- Experimental design is the “gold standard”
  - Randomization creates groups alike in all respects except for the presence or absence of the program
- Quasi-experiments and statistical controls seek to approximate randomization through use of natural variation
  - RAND/AIR NCLB evaluation

# Contrasted-Group Designs

## *Problems*

---

- However, contrasted group designs are often infeasible
    - Time and budget constraints
    - Political and ethical constraints in constructing no-treatment comparisons
    - Small samples/low statistical power
    - Low-intensity treatments
    - Variable treatments
-

# Theory-Based Evaluation

---

- Incorporates knowledge of drivers (causal mechanisms, in addition to raw empirical correlations)
- Use of logic models (example)
- See handout

# Theory-Based Evaluation

---

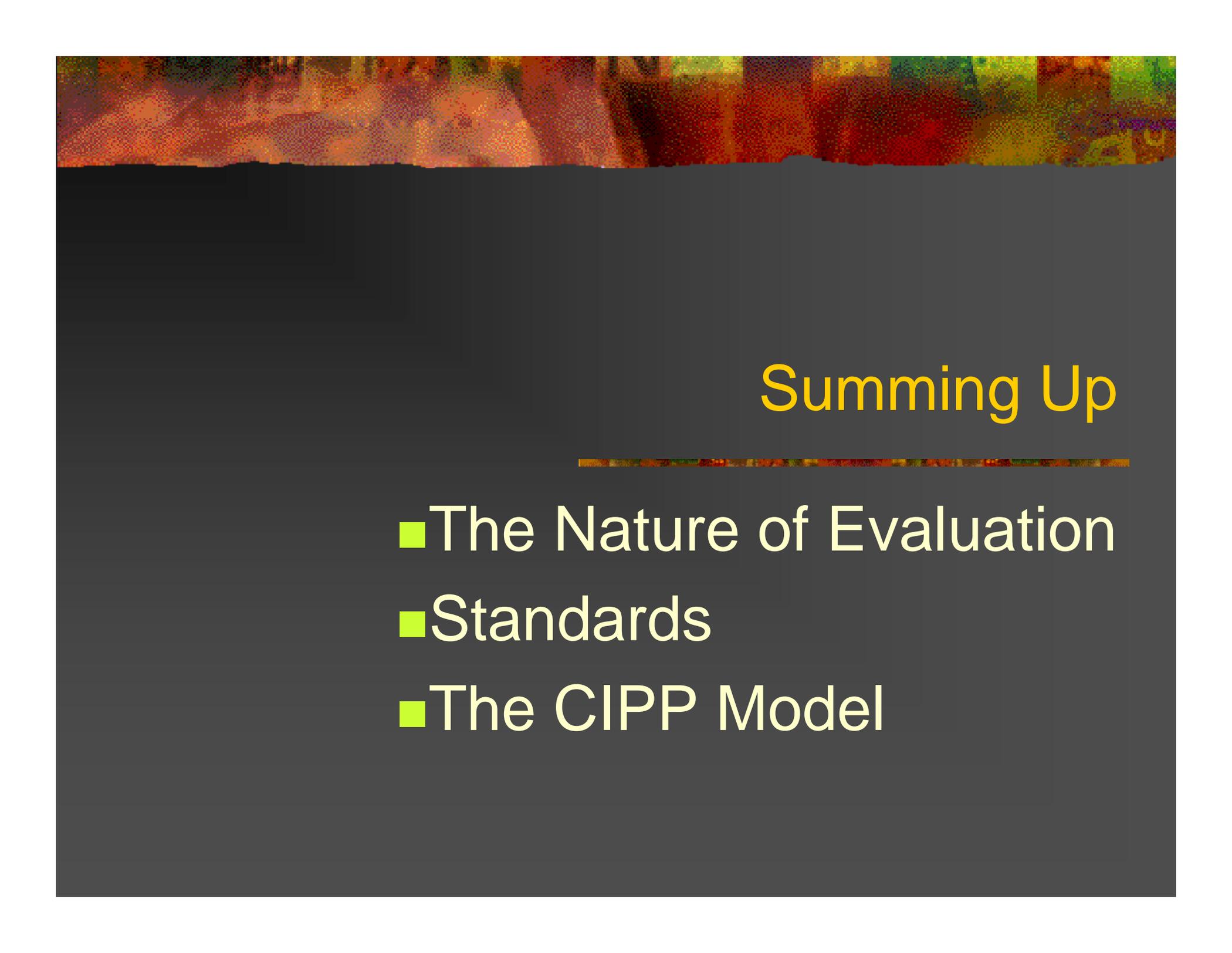
- Identification of leading indicators
  - Consideration of mechanisms can be useful in constructing recommendations
  - Useful in identifying non-obvious or unintended impacts
  - A complement, but not a substitute, for a good contrasted-groups evaluation
-



# Exercise Using Product (Impact) Evaluation

---

Chris Nelson



## Summing Up

---

- The Nature of Evaluation
- Standards
- The CIPP Model

# Summing Up

## *The Nature of Evaluation*

---

- Evaluation is the systematic determination of merit or worth
  - Thus, evaluation cannot be value free
  - Evaluation as a field has developed a strong stakeholder and utilization orientation
  - Formative evaluation seeks to apply evaluation at all stages of the program development cycle
-

# Summing Up

## *Standards*

---

- Program evaluation standards seek to draw together these ideas into a set of criteria for evaluating evaluations (meta-evaluation)
    - Utility
    - Feasibility
    - Propriety
    - Accuracy
-

# Summing Up

## *CIPP Model*

---

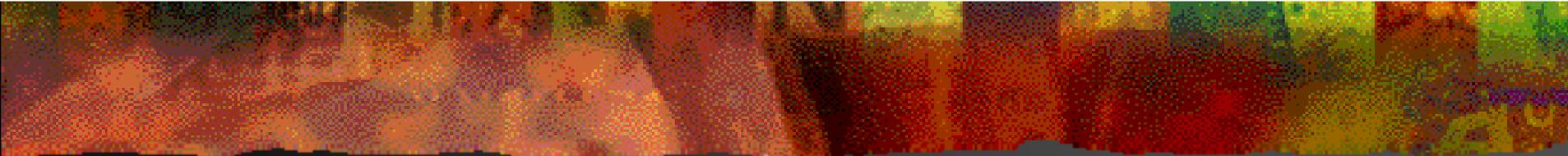
- CIPP model
    - Provides an inventory of evaluation questions and activities at all stages of the program development cycle (formative uses)
    - Can also be used summatively to improve usability of impact assessments
-

# The Evaluation Toolkit

## *CIPP Model*

---

- CIPP model (cont'd)
    - Context: What should we be doing?
    - Input: How should we be doing it/does the program match context?
    - Process: Are we doing it/did we do it?
    - Product: How effect is it?
-



# Looking Ahead

---

Building Your Own Evaluation  
Tool Kit

# Looking Ahead

## *Building your own evaluation toolkit*

---

- Questions? Discussion?
    - What is evaluation and how is it different from research?
    - Why is evaluation important in decision making?
    - How can evaluation be used at all stages of the program development cycle?
    - How can evaluations made more useful?
-

# Looking Ahead

*Building your own evaluation toolkit*

---

- What makes a good evaluation report?
- Scriven's Key Evaluation Checklist (KEC)

# Looking Ahead

*Building your own evaluation toolkit*

---

- Using evaluation checklists and other resources