Performance Measurement

Results-based Frameworks

Ann M. Doucette, PhD
The Evaluators Institute
The George Washington University
Washington, DC

Overview

- Performance Measurement Introduction
- Complexity – Systems Thinking
- Performance Measurement – Program Evaluation
- Program Logic
- Key Performance Indicators & Metrics
- Performance Measurement Frameworks

If you don't know where you are going, you might not get there.

Yogi Berra (2002). *When You Come to a Fork in the Road, Take It!: Inspiration and Wisdom from One of Baseball's Greatest Heroes*. Hyperion, p. 53.

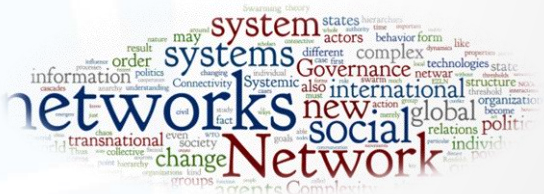
Conceptualizing Systems

Systems Thinking

	Easily knowable	+	Rules are given -- accepted	= Simple
Individuals and connections are equally important	+	Algorithms lead to predictable responses	+	Response is determined by the rules	= Complicated
Connections are essential; Individual representatives less critical	+	Simple rules lead to complex, adaptive, creative responses	+	Individuals have creative opportunity of responding within rules	= Complex

Simple

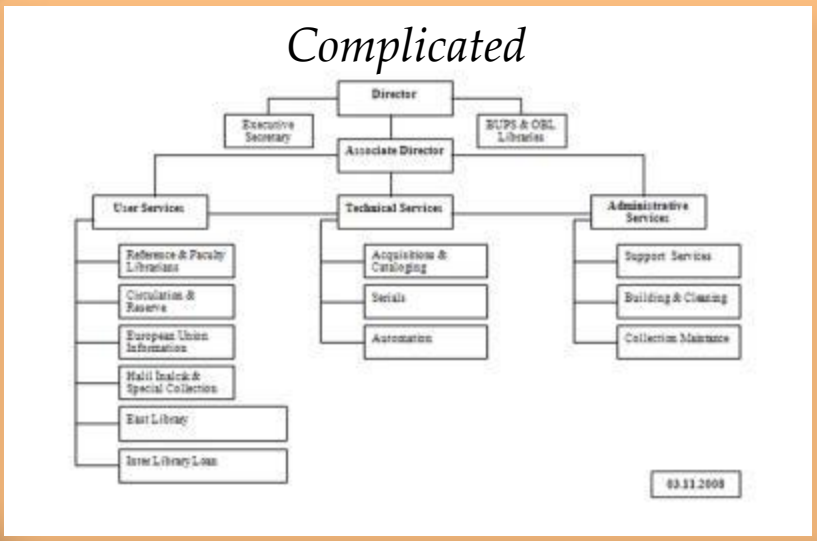
Complicated

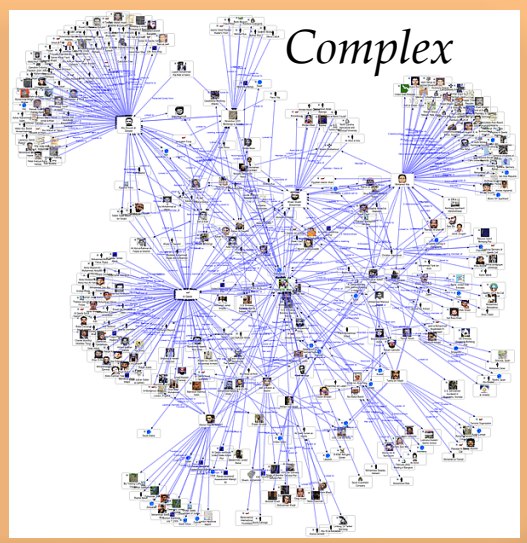


Complex

At times **CHAOTIC**

Complexity: Human & Social, Strategy & Operations, Technology – Economic Trends



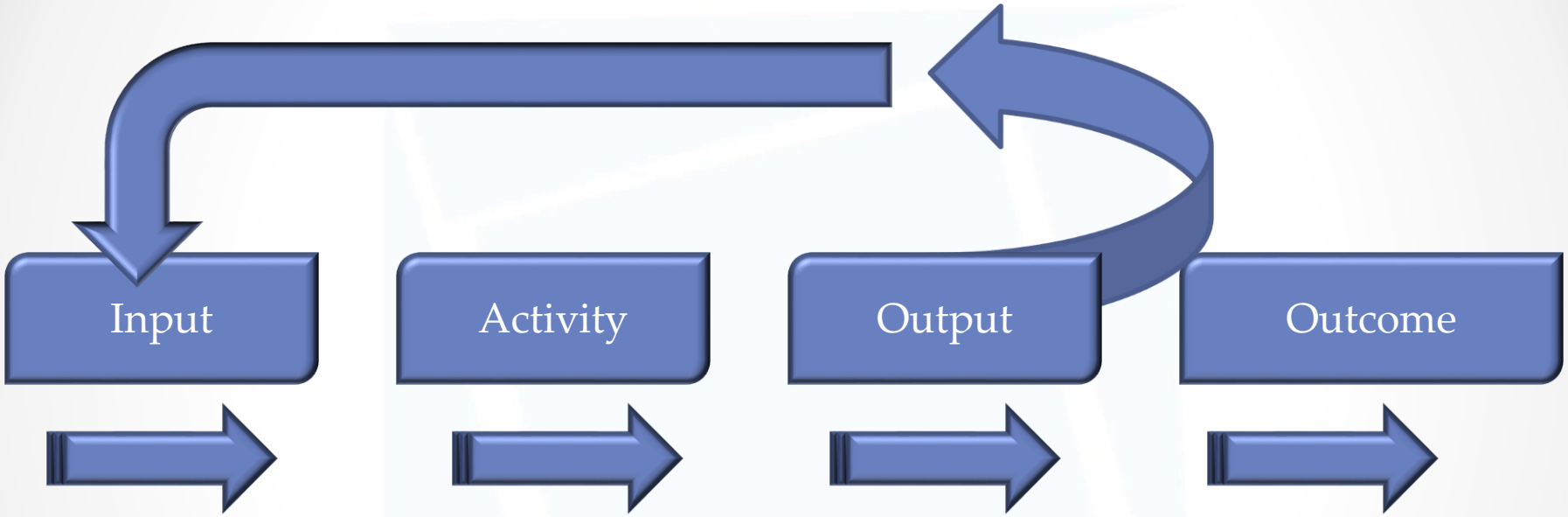


The product of partnership networks and associated interaction.

Traditional System Assumptions

- Linear – Cause → effect
- Outcomes can be understood through analysis
  - Identification of key components to explain change
    - Individuals: provider and patient/consumer characteristics
    - Treatment model (evidence-based, practice guidelines, etc.)
  - Study of the parts
- Analysis of past events → prediction of future outcomes

Traditional Approach: Linear Systems

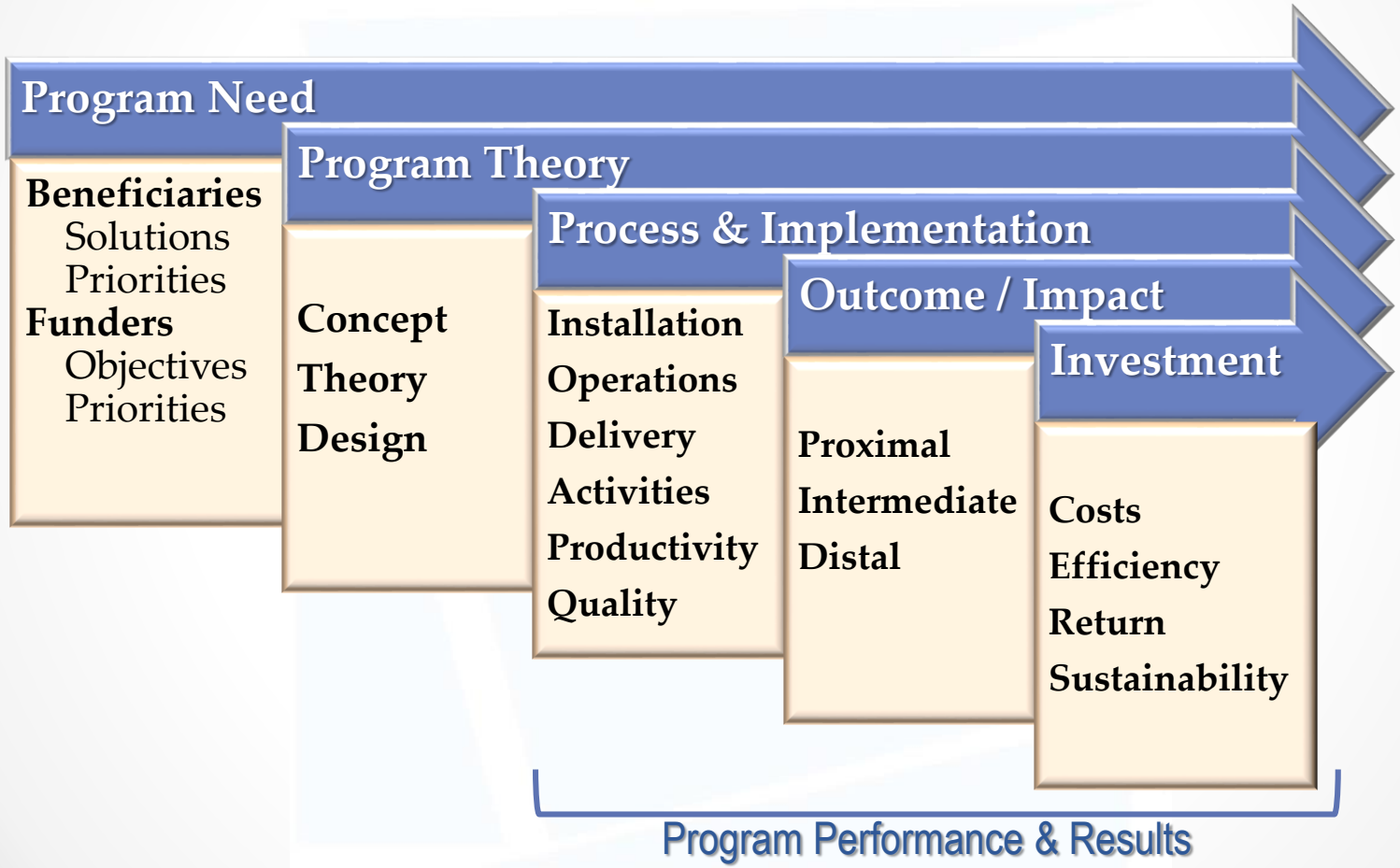


Complex Adaptive Systems → outputs become inputs in an iterative manner → optimizing program performance

Characteristics of Complexity

- Emergence – agents/representatives of the system interact in random ways
  - Interpersonal relationships
  - Coincidental meetings
- Connectivity – systems depend on interconnections and feedback
- Interdependence/co-evolution – systems operate within a specific environment (e.g., health, education, immigration, labor, etc.), however, they are also part of other systems (e.g., economic and policy environments)
  - Change in one environment precipitates the potential for change in another
- Rules – systems are governed by simple conventions
  - Example: Water systems (oceans, rivers, lakes, etc.) though different are governed by a simple rule – *water seeks its own level*
- Self-organization → consistencies and patterns
  - Example: Jazz – musical composition based on emergence (improvisation) and feedback
  - Change is often discontinuous
  - Outcomes may be different than anticipated

Program Evaluation Sequence



Results-based Frameworks

- Traditional: emphasis on tracking what is happening and what occurred
- Results-based: emphasis on pre-determined objectives
  - Monitoring: emphasizes progress toward achievement of objectives and goals
    - Identification of key indicators
    - Comparison between actual and expected results
      - Per-determined targets/results
      - Tracking evidence
        - Short-term
        - Intermediate
        - Long-term
  - Evaluation:
    - Is progress towards expected results associated with program?
    - Are there other explanations for results identified in the monitoring approach?
    - How effective and efficient is the program in achieving results?
    - Is this the most sustainable approach in achieving the expected results?

Program Evaluation
Performance Management

Program Evaluation

- To understand current program performance
- To describe program operations -- fidelity
- To assess effectiveness, efficiency and explain findings and results
- Focus: funders, policy –makers, other external groups
- Information: collected episodically
- Progress: assessed in terms of program objectives/goals
  - Rationale and understanding of performance
- Utility: Program decision-making

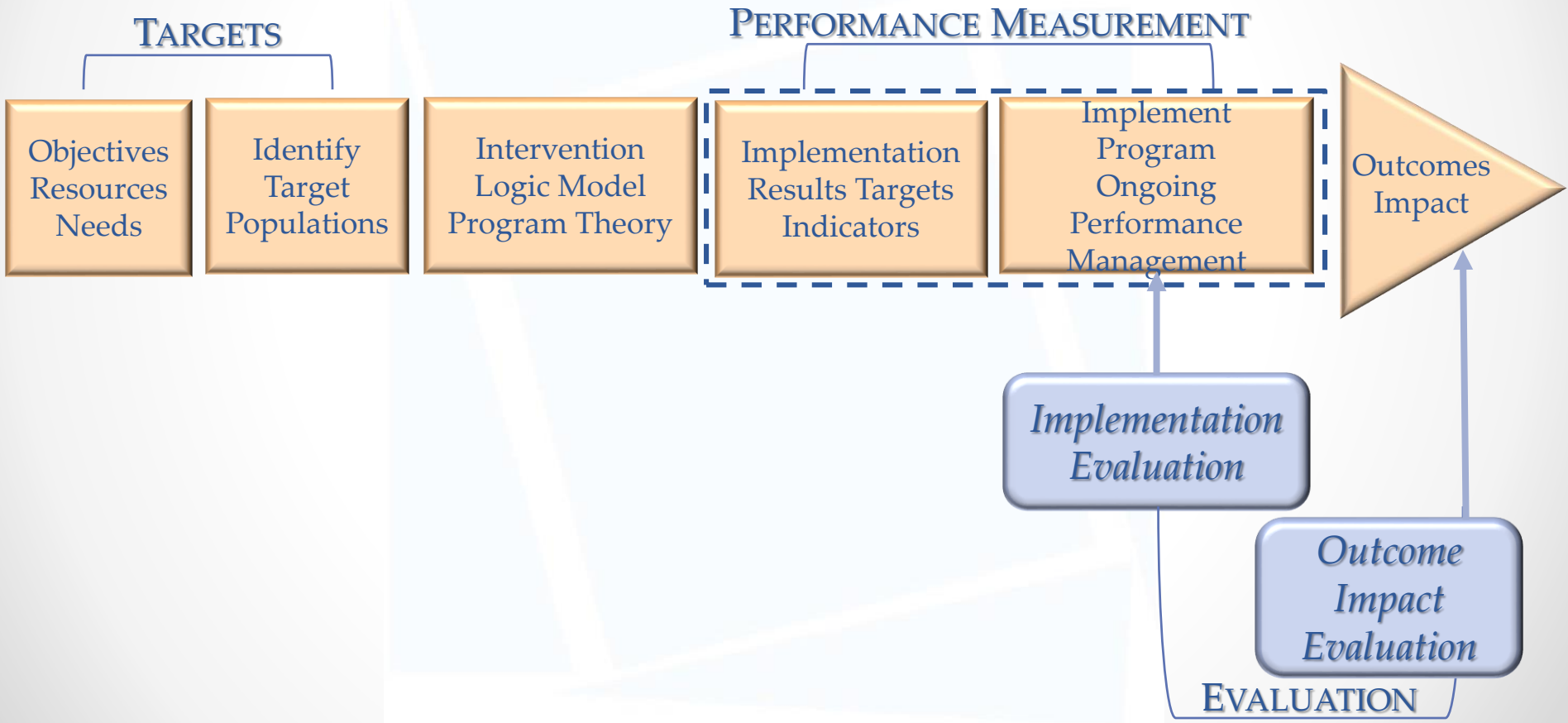
Performance Measurement

- To ensure program is functioning as intended
- To guide program improvement and optimization
- To examine the need for program modifications
- Focus: program managers and staff
- Information: collected throughout program cycle
- Progress: assessed in terms of increase and decrease in desired results
  - Depth of information: the *score*
- Utility: Continuous improvement

Performance – Definitions

- **Performance measures:** assessment approaches that provide organizations with a metric of program quality
- **Performance measurement:** process by which an organization monitors important aspects of its programs, systems, and processes
- **Performance management:** approach used to set goals and monitor progress toward achieving identified goals
  1. Organizations/agencies set goals
  2. Examine actual data using performance measures
  3. Acts on results to improve the performance toward its goals

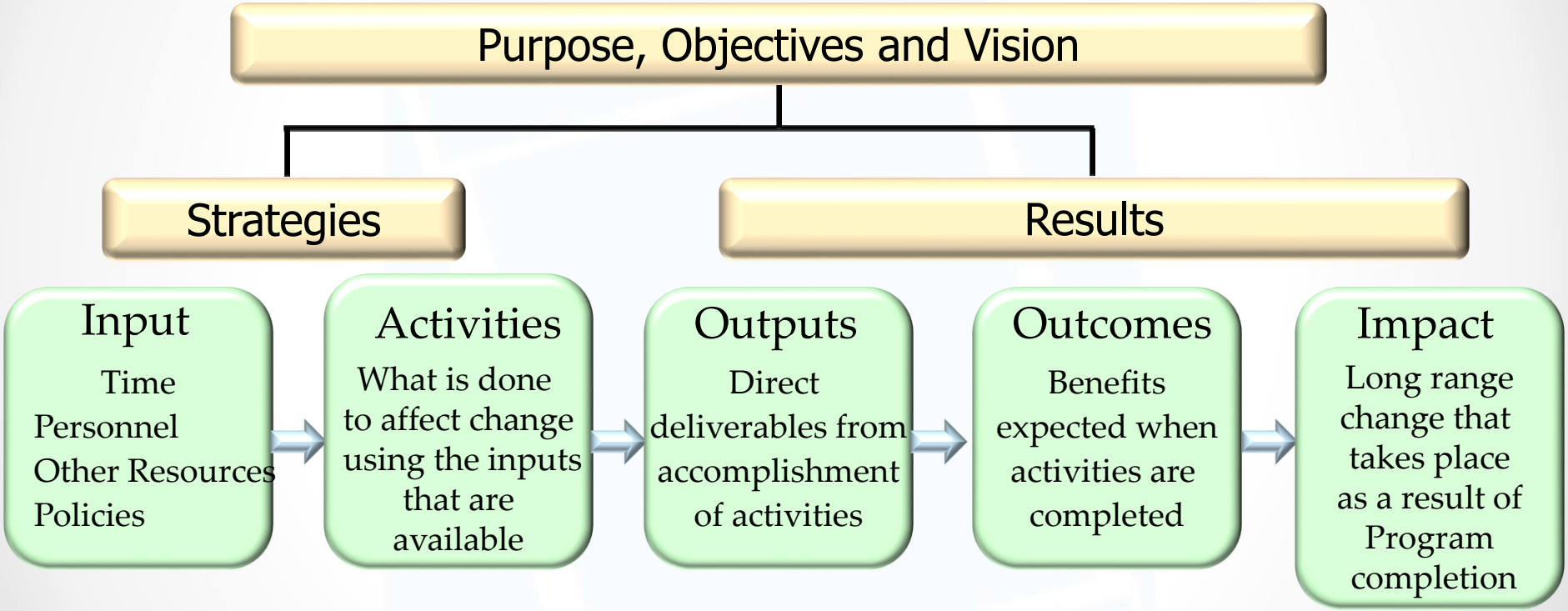
Performance Management Sequence



Developing a Logic Model

- Perspective
  - There are multiple points of view
    - Funders
    - Managers
    - Beneficiaries
    - Others
  - Who will be implementing the program?
  - Who will be managing the program?
- A logic model can reflect the perspective of
  - A single entity
    - Agency headquarters those commissioning the evaluation
    - Funders
    - Evaluators
    - Program staff
    - Others
  - A shared perspective across program partners

Logic Model


Purpose, Objectives and Vision
Strategies
Results
Input
Time
Personnel
Other Resources
Policies
Activities
What is done to affect change using the inputs that are available
Outputs
Direct deliverables from accomplishment of activities
Outcomes
Benefits expected when activities are completed
Impact
Long range change that takes place as a result of Program completion

Outputs – Outcomes

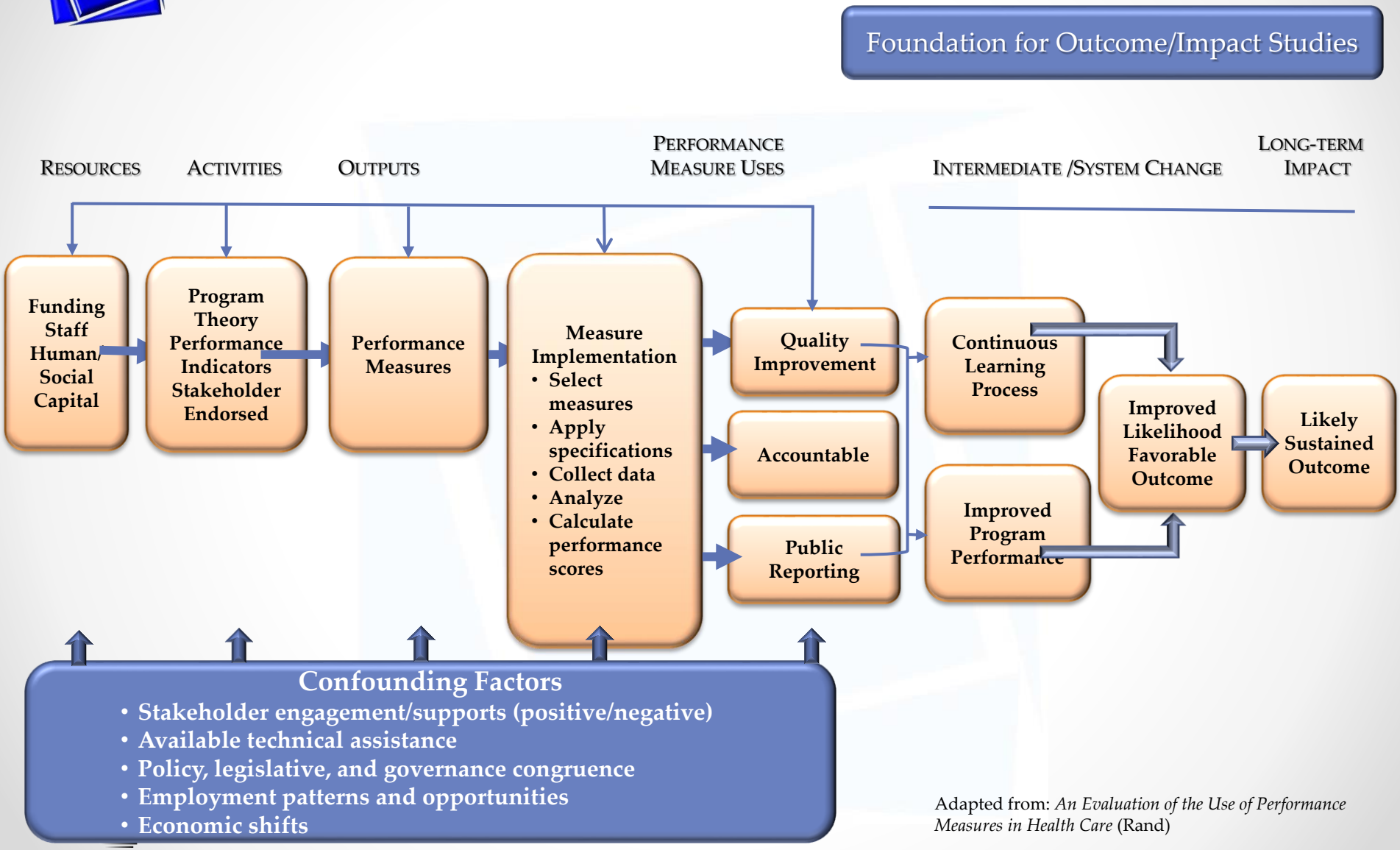
Outputs

- What will be done to achieve results
- Frequency, intensity, duration of program activities and/or participation
- Understanding of outputs is needed to replicate results

Outcomes

- What beneficiaries will gain from the program
- Change that occurs (beneficiaries, communities, organizations, etc.
- Program contributions to results
- Evidence for program decision-making

Performance Measurement Logic Model



Adapted from: *An Evaluation of the Use of Performance Measures in Health Care* (Rand)

Using the Theory of Change Framework

- Theory of change is a framework against which the success and obstacles of your initiative will be evaluated
- Assumptions, justification, and contextual conditions believed to affect
  - Program/initiative success, information about how assumptions change, expand, or prove correct can be evaluated against the initial theory of change and modification and adaptations made throughout the program/initiative

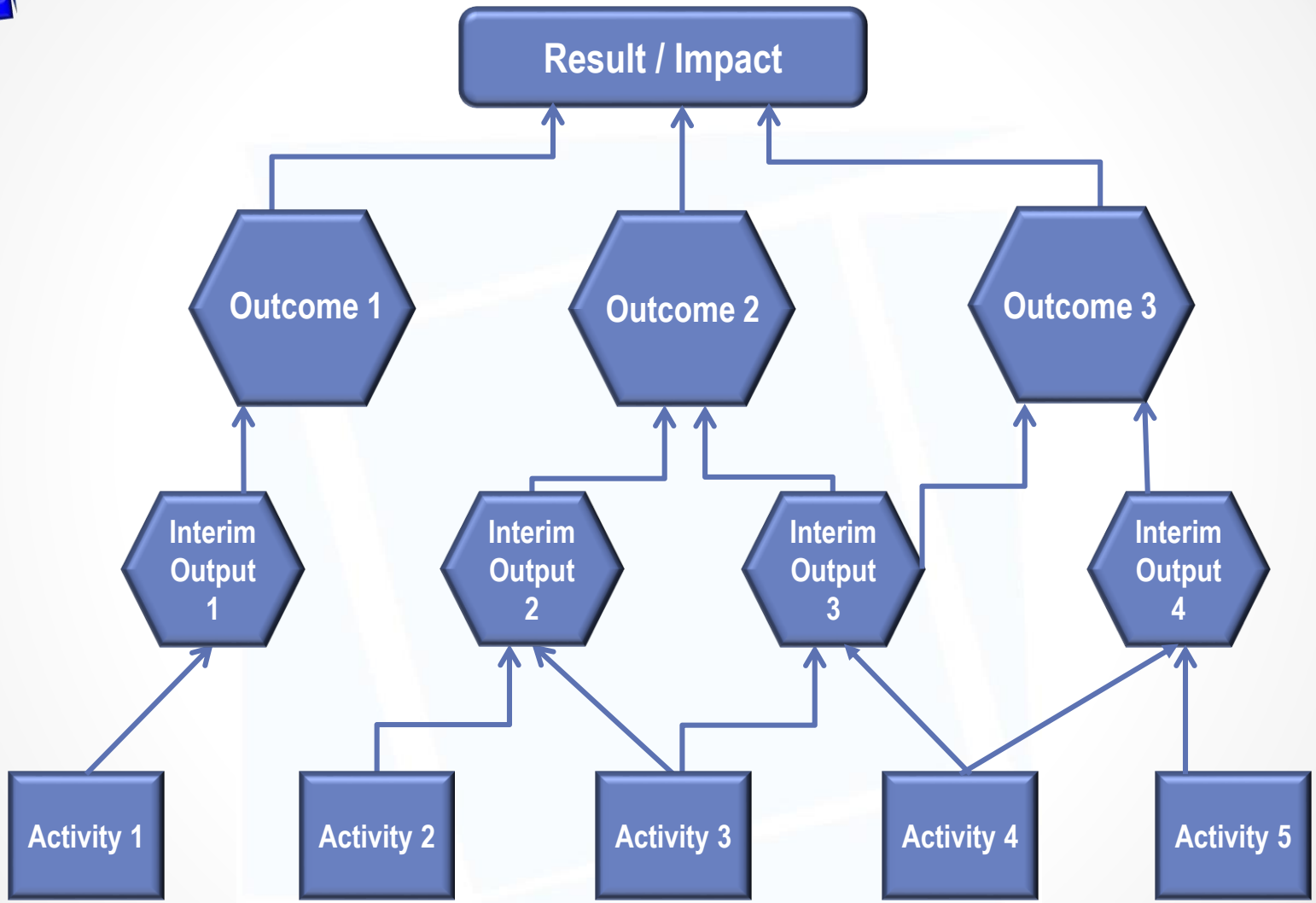
Steps in Building a *Theory of Change*

1. Identifying long-term goals and the assumptions behind them
   a) Identification of connections between long term, intermediate and early outcomes
   b) Verification that the important preconditions for success have been identified
   c) Justification that the links/mechanisms between program activities and the outcomes are as expected
   d) Identification of contextual (political, economic, etc.) and environmental factors that will support or hinder progress

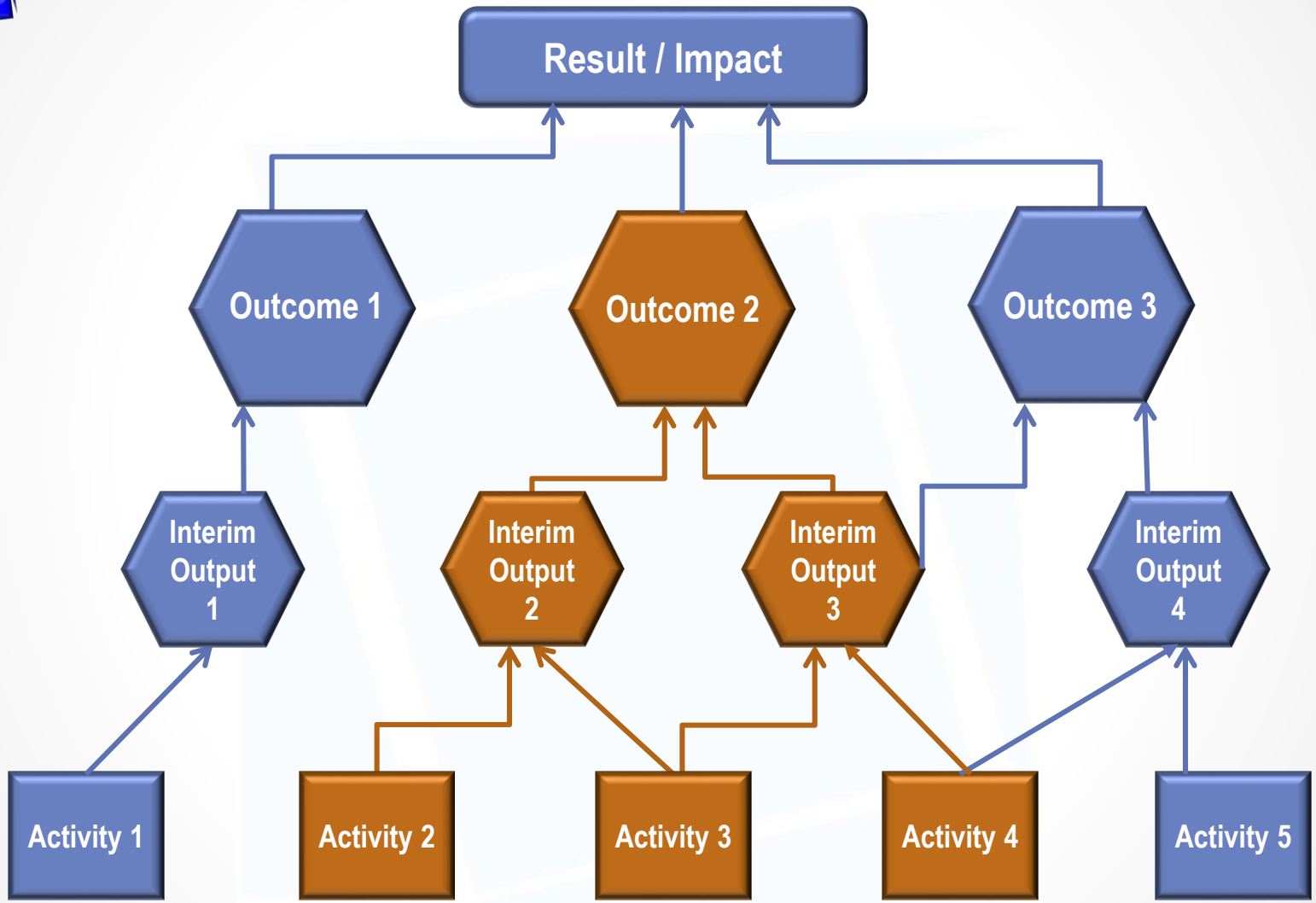
Steps in Building a *Theory of Change* *(cont.)*

2. Backwards mapping and connect the preconditions or requirements necessary to achieve that goal
3. Identifying the interventions that your initiative will perform to create your desired change
4. Developing indicators to measure your outcomes to assess the performance of your initiative
5. Writing a narrative to explain the logic of your initiative

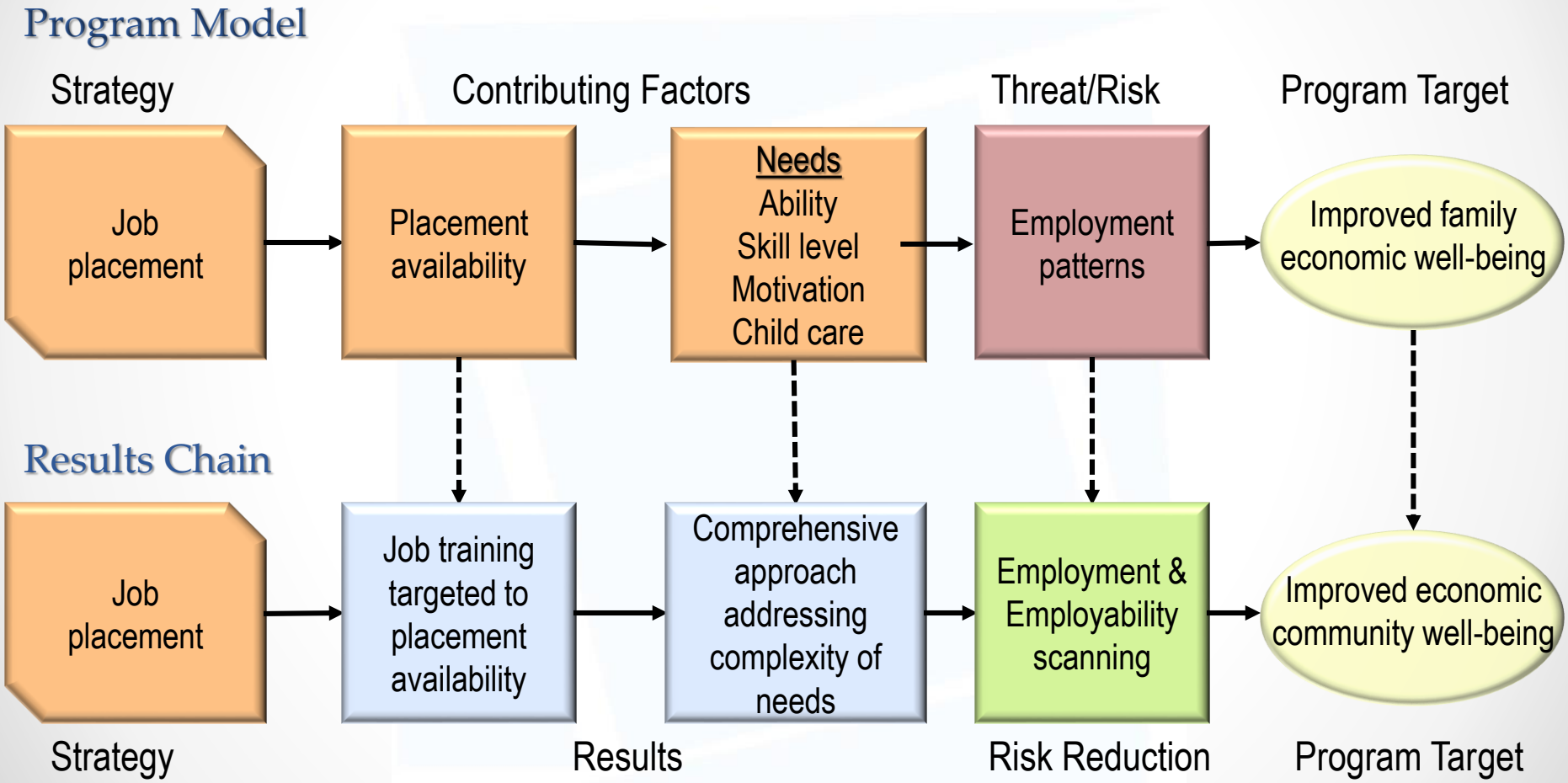
Theory of Change: Multiple Pathways



Theory of Change: Multiple Pathways



Employment – Family Stability



Adapted from Margoluis, R. et al. (2013). Results Chains: a Tool for Conservation Action Design, Management, and Evaluation. Ecology and Science 18(2)

Performance Indicators

Performance Measurement Systems

- Clearly defined measures
  - Captures program performance and interests of stakeholders
- Readily understood by stakeholders
- Directly related to planning objectives
  - Translates program performance into measureable metrics
- Relevant to decision-makers and stakeholders
- Reflects reliable program performance over specified duration
- Sufficiently addresses risk and uncertainty
  - Ability of performance measurement system to adapt to changing project conditions
- Adaptive measurement approaches to capture program changes

Constructing Indicators

- Indicator development should be informed by multiple stakeholders
  - Technical expertise
  - Local knowledge
  - Policy and decision-makers
- Indicators should provide continuous feedback throughout the program, policy cycle, etc.
- Indicators should monitor
  - Inputs
  - Activities
  - Outputs
  - Outcomes
  - Indicators should provide alerts
    - Underperformance in reaching targets
    - Variability in performance
- Indicators should be specific to desired outcomes

Direct Measures – Indices – Scales

- Direct measure: comparison of direct measure with known standard, no mathematical calculation needed
  - Gender Equality – can it be assessed using a direct measure?
- Index: composite measure – accumulation of scores across multiple variables representing ranked ordered observations
  - Gender Equality Index: *Level of Civic Engagement*
    - Ability to vote, % women registered to vote, % of women voting, number of women running for municipal positions, # women elected to municipal positions)
- Scale: composition of scores (mean, sum) across variables having an empirical structure
  - Gender Equality Scale: *Perception of prejudice against women regarding civic engagement*

Indices versus Scales

Similarities

- Ordinal assessments
  - Rank
  - Subjective – strongly agree to strongly disagree
- Composite measures
  - Data across multiple variables contribute to overall score

Differences

- Index: items/variables are additive
  - Index items may or may not be intercorrelated; can be weighted
  - Example: index of adolescent risk: tobacco use, drug and alcohol use, skipping school, conflict with peers, destruction of property
- Scale: items/variables are intercorrelated, items can be weighted, contributing more to the scale score than do others (Likert scale)
  - Example: depression: feeling sad, hopeless, no purpose in life, worthless

The S.M.A.R.T.* Approach

- Specific: precisely defined objectives
- Measureable: progress toward objectives can be measured
- Achievable: expectations that indicators are able to be tracked and will change given the environment and existing resources and time
- Relevant: program outcomes support organizational mission and goals and be realistic and feasible to collect
- Timely: clear statements of when objectives will be accomplished, knowledge of time taken to collect data, lag between activities, outputs and outcomes

S.M.A.R.T.**E.R.**

- Evaluative culture: environment where stakeholders are curious about results
- Routine : M&E is a routine component of programmes

*Drucker, 1954; Krick, et al., 2005

Developing Indicators – Avoiding Ambiguity

- What is meant by . . .
  - Timely case processing
  - Job placement
  - Obtaining employment
  - Earning gain
  - Increased educational attainment
  - Self-sufficiency
  - Family stability
  - Publication of results

Developing Good Indicators

Indicators are management and monitoring tools.

- Action: does knowing information from the indicator assist program and/or stakeholders in doing things more effectively/efficiently?
- Is the indicator relevant to the program? It is important to the stakeholders?
- Does the indicator reflect change – the effects the program has had?
- Is the indicator information understandable by all stakeholders?

Proposed Criteria for Indicator Development*

1. Clear and straightforward: simple to compile and interpret
   i. Composite indicators should be avoided where possible since they require more complex data collection methods
   ii. rely on imputation for missing variables, and arbitrary weighting
   iii. Difficulty in informing policy recommendations.
2. Consensus based in line with existing standards: Indicators should be supported by a consensus on measurement
3. Indicators should draw on well-established sources of public and private data and be consistent to enable measurement over time

*Adapted from United Nations – Proposed Sustainable Development Goals

Criteria for Proposed SDG Indicators

4. Indicators should draw on well-established sources of public and private data and be consistent to enable measurement over time
5. Indicators should lend that lend themselves to disaggregation according to
   i. Characteristics of the individual or household (e.g. gender, age, income, education, disability, race/ethnicity)
   ii. Geographic dimensions (e.g. by metropolitan areas, urban and rural, or districts).
6. Universal: The set of indicators as a whole needs to track the Family Stability/TANF/WIA agenda. Many (though not all) of the indicators should be applicable across states

Criteria for Proposed SDG Indicators

7. Indicators should emphasize outcomes, but not ignore inputs and outputs
   i. Choice between input and outcome measures must be handled pragmatically
   ii. Input metrics can play a critical role in driving and tracking the changes needed for a successful outcome
   iii. For example, access to childcare services is a vital component of family employment. Dedicated indicators are needed to track both (access and availability of such services)

Indicator Examples

a) Primary school completion rates for girls and boys

or

b) Percentage of girls and boys who master a broad range of foundational skills, including in literacy and mathematics by the end of the primary school cycle (based on credibly established national benchmarks)

---

a) Gini coefficient – income dispersion (inequality)

$$G = \sum_{N-1}^{N+1} - \sum_{N(N-1)\mu}^{2} \left(\sum_{i=1}^{n} P_i X_i \right)$$

- μ= mean population income
- P_i income rank of individual with X_i income

or

b) The bottom 10% of the population has .2% of the national income, while the top 10% has 63% of the national income

Indicator Criteria

- Do the indicators reflect stakeholder and/or beneficiary needs?
- Do the indicators address important program performance dimensions?
- Are there identified sources of credible data?
- Are indicator results understandable?
- If the cost (labor, funds) of the indicator data feasible and practical?
- Does movement on the indicator reflect favorable and/or unfavorable results?
- Can the indicator be sustained over time?

Performance Targets

Indicator Targets

- Specific
- Operational
- Quantifiable
- Attainable
  - Some *stretch goals*
- Applicable to stakeholders
- Transparent – addresses accountability
- Adaptable for local contexts
  - Starting points may differ across regions
- Clear definition on meaning of universal and zero
- Direct measures if available (scales and indices otherwise)

Targeting Improvement

- How much progress can be made?
  - In how much time?
- What is the starting point – baseline?
- How is progress perceived in terms of the theory of change?
- Are there sufficient resources distributed across the program timeframe?
- Will additional resources beyond the program be needed?
- What other factors might contribute to the success or failure of the program?
- Political concerns?
- What is the history of the organization(s) responsible for delivering the program?

Performance Targets

- Performance targets focus the program on what is to be accomplished
  - Clear management expectations
- Final and interim targets
  - Expected program achievement
  - Interim goals → desired outcome
    - Changes in relationship to the scale of the problem
    - Development and sustainability of novel approach
    - Quality
    - Error rates (reduction)
    - Efficiency

Performance Target Considerations

- Performance baseline – *starting point*
- Historical trends – indicator(s) function prior to the program – patterns
- Stakeholder expectations of progress
  - Implementing partners
  - Beneficiaries
  - Funders
  - Expert judgment
- Progress is not always linear

Benchmarking – Setting Targets

. . . Identifying the highest standard of performance, articulating the processes taken to achieve that standard, modifying and applying the processes for continued improvement

Benchmarking

- Strategy – what strategies are used to achieve high performance
  - Competitive
  - Targets identified to optimize performance and quality scores
    - Focus is on improvement in scores/outcomes
- Process
  - Procedures used to . . .
    - Identify need
    - Access intervention/programs
    - Exchange necessary information
    - Coordinate efforts
- Best Practices
  - Use of identified standards
  - Focus on processes and the management activities supporting them
- Cost – resource use

Ann M. Doucette, PhD

Benchmarking 101

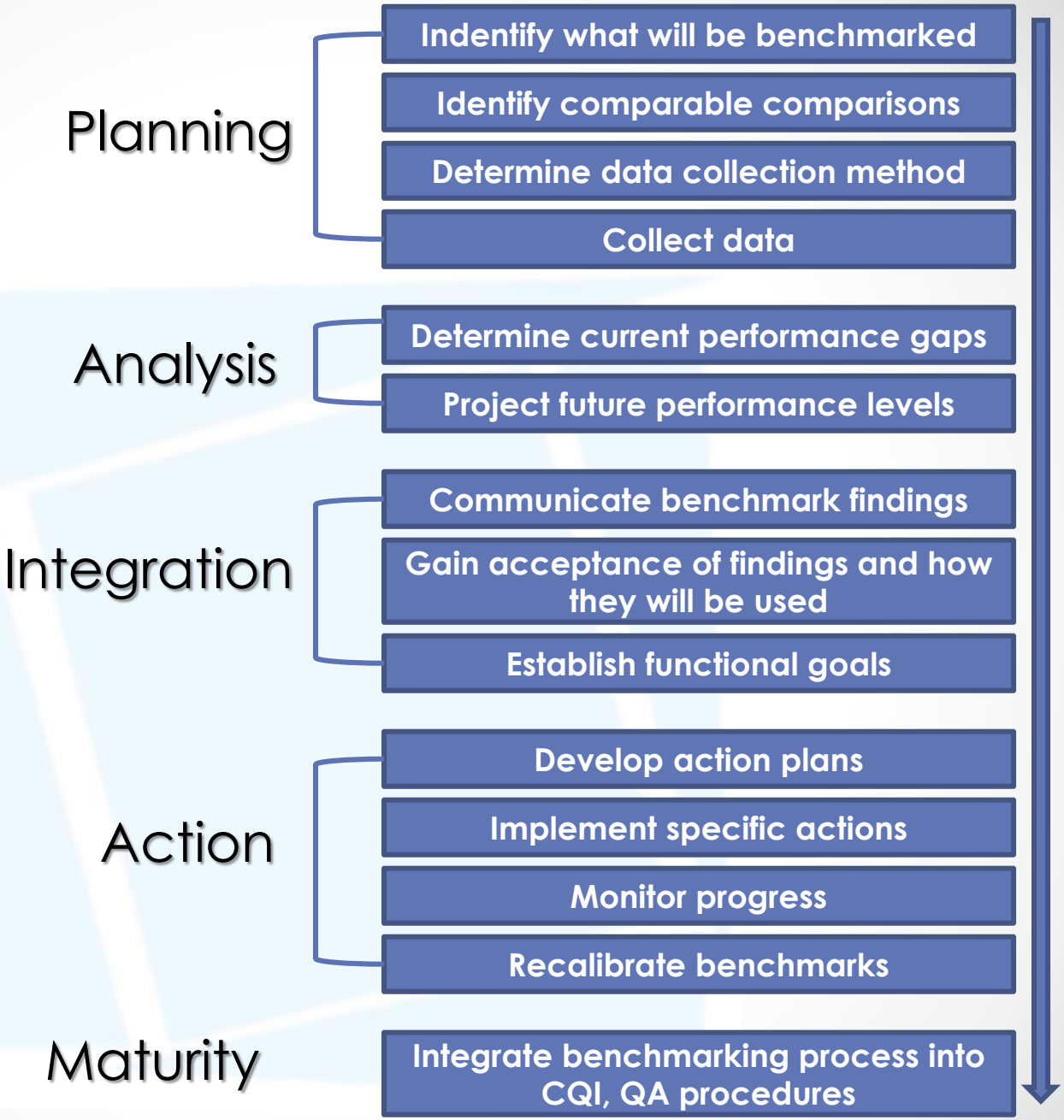
- Benchmark: reference value for identified indicators
  - External: comparison to peer organization/agency
  - Internal: self-comparison – present performance compared to past performance
- Indicator: data (numbers) that reflect the measurement of performance activity
  - Completion rates
  - Utilization
  - Impact
- Peer group: organizations/agencies with similar characteristics that are relevant to benchmarking performance

Ann M. Doucette, PhD

Benchmarking 101 (cont.)

- Benchmarking is not a ranking, rather it establishes levels of optimal improvement – *target goals*
  - Comparison and references with what other systems have achieved – *best practices* is useful, **but is not benchmarking**
    - Benchmarking is the distance from where the system is currently to where the system wants to be – *performance gaps*
- Steps to establish benchmarks
  - Identify what to benchmark
  - Select key performance indicators
  - Identify benchmarking partners– objective assessment of progress
  - Establish data collection approach (quantitative & qualitative)

Ann M. Doucette, PhD



Adapted from: Camp, R. C., (1998). *Benchmarking: The search for industry best practices that lead to superior performance*. ASQ Press: Milwaukee, WI.

Benchmarking Steps

1. Review the literature
2. Interview stakeholders
3. Establish performance objectives
4. Develop – *brainstorm* list of indicators
   a) Review for redundancy
   b) Prioritize indicators in terms of
      i. Relevance
      ii. Feasibility
5. Define *candidate* indicators
6. Operationalize indicators
   a) Definitions (inclusion and exclusion criteria)
   b) Scaling
   c) Timeframe
7. Achieve consensus across stakeholder groups on performance standards
   a) What performance aspects are most valued by various stakeholder groups?
8. Validation
9. Establish baseline
10. Develop implementation plan

Caution In Using Benchmark Metrics

- Organization and agencies should be stable across the benchmarking interval
  - Instability may erroneously be interpreted as improvement or deterioration as a consequence of system flux
- Random variation related to sample size must be considered
  - The amount of random variation is inversely related to sample size
    - Variation is more noticeable in small samples, small agencies may be characterized as atypical as a result of sampling error
- Consider special characteristics of organizations/agencies
  - Difference from the benchmark target does not automatically indicate the need for improvement
- Consider the potential for unintended consequences
  - Staff morale
  - Team work – if benchmarking is seen as competition team work may deteriorate
- Learning from the performance of others

Ann M. Doucette, PhD

Evaluation Shortcoming (examples)

Performance measurement cannot . . .

- Identify all possible contributions to program progress
  - Functional improvement may be linked to the program, community resilience, social-psychological variables, socio-economic conditions, cultural norms, etc.
  - If these are not measured no attribution of causality can be made.
- Assure quality
  - Quotas (outputs), such as attendance at program workshops say nothing about the quality of instruction or application of learning and change in behavior.
- Capture the entire system
  - Measures reflect only those program components assessed.
  - Completed data, especially follow-up data with substantial attrition may not be generalizable in terms of speaking to program effectiveness or efficiency.

Challenges: Quality and Performance Monitoring

The most common mistake organizations make is measuring too many variables. The next most common mistake is measuring too few.

Mark Graham Brown
Keeping Score (1996)

. . . But perhaps an even bigger mistake is to keeping doing the same thing, hoping for better results.

Ann M. Doucette, PhD

Thank You!

Ann Doucette, Ph.D.

The Evaluators' Institute

MIDGE SMITH CENTER FOR EVALUATION EFFECTIVENESS

The George Washington University
2036 H Street NW, Suite 105
Washington, DC 20052
Tele: 202.994-8112
Email: doucette@gwu.edu
Website: www.tei.gwu.edu