

AEA/CDC Summer Evaluation Institute

Offering 7: Using Program Evaluation to Improve Nonprofit Outcomes

Description: Funders, communities, and service recipients themselves ask that nonprofits show results and demonstrate progress towards service goals. Yet, identifying, measuring, and reporting on service outcomes in ways that attend to the multiple stakeholders involved can be a challenging task. This workshop will help evaluators, program officers, and program administrators working in the nonprofit sector by providing an overview of the practice of outcomes assessment tailored to the nonprofit context. Attendees at the session will:

- Learn what/who drives program evaluation and outcomes assessment in nonprofit service providers
- Explore uses of performance data in the nonprofit sector
- Discuss challenges to measurement of social service program outcomes, and
- Develop strategies for making measurement decisions.

Audience: Evaluators, program officers, and program administrators working in the nonprofit sector.

Kathryn Newcomer, Ph.D. is the Director of the doctoral program in Public Policy and Administration, and Associate Director of the Trachtenberg School of Public Policy and Public Administration, at the George Washington University where she teaches program evaluation, research design, and applied statistics. She is a sought-after trainer on the topics of performance measurement and program evaluation and has authored or co-authored numerous journal articles as well as five books, among them: *The Handbook of Practical Program Evaluation* (1994, 2004), and *Meeting the Challenges of Performance-Oriented Government* (2002). She was identified as one of the top 25 evaluation experts in the country in 2001 by the *American Journal of Evaluation*, is a Fellow of the National Academy of Public Administration, and currently serves on the Comptroller General's Educators' Advisory Panel. Dr. Newcomer earned her Ph.D. in political science from the University of Iowa.

Offered (Two Rotations of the Same Content - Do not register for both):

- Monday, June 23, 9:25 – 12:45 (20 minute break within)
- Tuesday, June 24, 9:25 – 12:45 (20 minute break within)



Using Program Evaluation to Improve Nonprofit Outcomes

A Workshop For

AEA/CDC 2008 Summer Institute

**Kathryn E. Newcomer, Ph.D
The School of Public Policy and Public Administration
The George Washington University
June 2008**

1



Session Objectives

- Learn what/who drives program evaluation and performance measurement in public and nonprofit service providers
- Explore uses of evaluation and outcomes assessment in the non-profit sector
- Discuss challenges to measurement of social service outcomes
- Clarify where to start in using evaluation to improve nonprofit social service programs, and
- Develop strategies for making measurement and reporting decisions

2



First things first...

Program Evaluation is: The application of systematic analytical (social science research) methods to address questions about program operations and results; and

Performance Measurement is: The routine measurement of program inputs, outputs, intermediate outcomes or longer-term outcomes attributed to a program; and

Both involve measurement plus judgment!!

3



Distinctions in Evaluation Approaches and Foci

Formative-----Summative

Ongoing (PM)-----One-Shot

Objective Observers----- Participatory

Quantitative Data-----Qualitative Data

Problem Orientation-----Non-problem

4

Theory Underlying Program Evaluation Practice



- Evaluation and/or programmatic performance measurement of programs is undertaken in order to improve programs and their outcomes -- through providing useful and timely information about programs.
- Performance data may be collected routinely, such as annually, and can address questions about quantity and perceived quality, e.g., “how many clients were served”, and “how satisfied were they with the services?”
- A more extended effort to evaluate programs is need to answer “Why” and “How” and “With what result” questions about services
- Both performance data and evaluation studies undertaken can be useful in the exercise of accountability

5

What/who drives program evaluation and performance measurement in public and nonprofit service providers?



6

Forces Encouraging Program Evaluation Among Nonprofits



- Government
 - The Government Performance and Results Act and OMB's PART process at the federal level
 - "Managing for Results" initiatives in states and cities
- United Way
- Foundations
- Boards of directors of nonprofits
- Professionalized staff
- Evidence-Based Policy Movement
- Other Donors

7

How is evaluation (and performance measurement) used in the non-profit sector?



8

Purpose of Program Evaluation and Performance Measurement in Nonprofits



- Understand and improve programs
- Test theories of change/program theories
- Tell a program's story
- Be accountable
- Inform the field
- Support fundraising efforts

9

Reported Use of Outcomes Data by Social Service Providers (Urban Institute, 2002): Internal Uses



- Clarifying Agency and Program Purposes for Staff
- Identifying Where Improvement is Needed
- Identifying What Works and Good Practices
- Informing, Training, and Motivating Staff
- Seeking Explanations and Outcomes
- Identifying Trends and Making Other Comparisons
- Using Outcome Data to Help Motivate Clients
- Giving More Voice to Clients
- Supporting Budgeting and Planning
- Facilitating In-Depth Studies and Program Evaluation
- Informing the Board of Directors

10

Reported Use of Outcomes Data by Social Service Providers (Urban Institute, 2002): External Uses



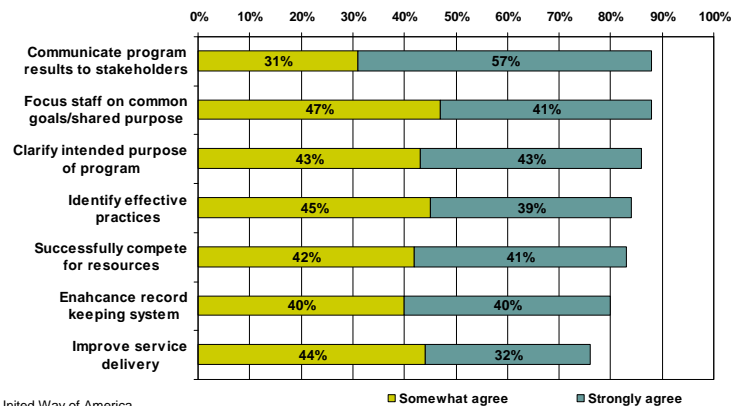
- Increasing Accountability to Funders and the Community
- Providing Data for Marketing and Fundraising
- Attracting Volunteers and Clients
- Helping Other Organizations

11

How Nonprofits Say Evaluation Helps



"Implementing outcome measurement has helped us to..."

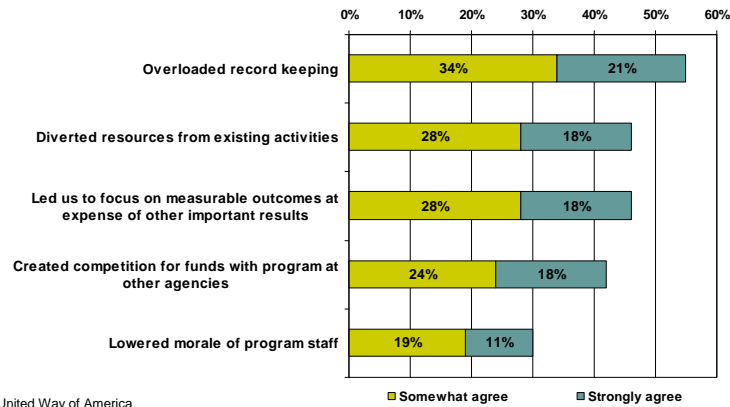


12

How Nonprofits Say Evaluation Hurts



"Implementing outcome measurement has..."



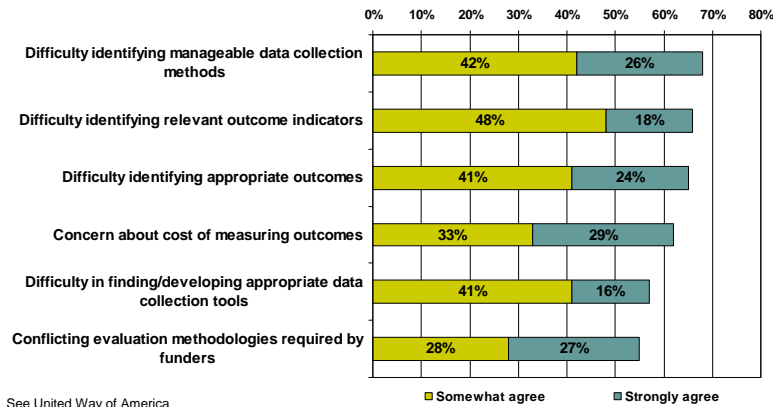
See United Way of America

13

Evaluation Obstacles in Nonprofits



Presence of barriers to measuring program outcomes included:



See United Way of America

14



What are challenges to measurement of social service outcomes?

15



Challenges to Measuring Performance in Social Service Delivery: WHY??

1. What is the espoused use of data?

Assess Program's Worth-----Improve Programs

2. Who is requesting the evaluation data?

External Funders-----Internal Staff

3. How might the evaluation data be used?

Empower Program Detractors----- Empower Program Managers

16

Challenges to Measuring Performance in Social Service Delivery: WHAT??



4. What will be measured?

Controllable Outputs-----Longer-term Outcomes

5. How many things will be measured?

Insufficient/Incomplete Data-----Information Overload

6. How clear and testable is the theory of change underlying the program(s)?

Complex System----- Clear Pathways

17

Challenges to Measuring Performance in Social Service Delivery: WHO??



7. Where does responsibility for data collection fall?

Evaluation Staff-----Program Staff

8. Who will bear the burden of data collection?

External Evaluators-----Frontline Staff

18

Challenges to Measuring Performance in Social Service Delivery: WHEN??



9. When will we measure?

Six Months or Longer-----Immediately

10. When will we report?

Annually or less frequently-----Quarterly or more often

19

Challenges to Measuring Performance in Social Service Delivery: HOW??



11. Where are the resources for evaluation?

Operating Program Resources-----Evaluation Set-asides

12. Who will participate in evaluation processes?

Funders -----Internal, Programmatic
Personnel, Leadership, and
Volunteers

13. How will participants' contributions count?

Scientific (Objective Evaluation Expertise Trumps)-----Democratic (All Stakeholders' Views Deemed Equal)

20

Challenges to Measuring Performance in Social Service Delivery: HOW??



14. How is evaluation institutionalized?

Infrequent/Add-on to Workload ----- Ongoing Management Function

15. How consistent are internal incentives supporting evaluation?

Very Little/Inconsistent Support ----- Consistent Leadership Support

16. To what extent will the validity and reliability of the data be ensured?

Not at all ----- On a Regular Basis

21

Challenges to Measuring Performance in Social Service Delivery: HOW??



17. How conducive is the organizational culture to evaluation?

in terms of:

Clarity in Vertical and Horizontal Internal Communications
Workforce Stability
Level of comfort with Quantitative Analyses
Receptivity to Organizational learning
Support for Risk-taking

22



Where do you start in using evaluation to improve nonprofit social service programs?

23

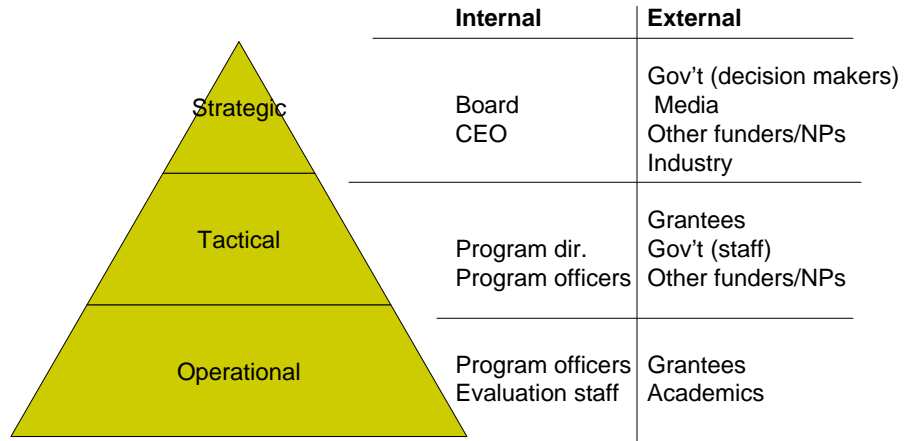


Strategies for Using Evaluation to Strengthen Nonprofit Programs: First Considerations

- Audience for data?
- Use of data?
 - Balancing use for accountability versus program improvement?
- Clarity of theory underlying program?

24

Foundation Evaluation Audience & Needs



See Baxter, p. 287

25

What are the implications of the different types of use: internal versus external, and strategic, tactical and operational?



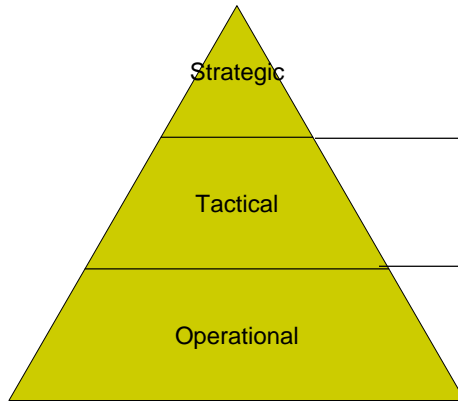
- Strategic?
- Tactical?
- Operational?
- And at each level – external or internal users??

26

Implications for Information Needs?



Characteristics of Data Needed?



Fewer metrics, global, trends, comparative, anecdotal success stories, and ??

All of above, programmatic outcomes linked to budgets, and??

All of above, and disaggregated across client characteristics, and ??

27

Implications for Reporting?



Less



More

28

What difference does it make whether the information or data are used internally or externally?



- Can the same data inform both external accountability requirements and management decision-making?
- What are common games played in using data for accountability?

29

Potential Consequences of Performance Reporting for External Audiences?



- “Do you count what can be counted rather than what counts?” (Einstein)
- Is the rush to measure expanding our capacity or is capacity shaping measurement?
 - Are we adequately auditing validity and reliability of data?
- Are we interpreting the numbers out of context? (any systems thinking?)
- Are calls for “hard evidence of effectiveness” in some areas even within reason?
- What is the impact of setting targets?
 - Threshold effects?
 - Outputs distortion?
- What about rankings?
 - Validity of criteria?
 - Reliability of data used?

30

What is needed to produce useful information about program performance?



- An understanding of how a program is supposed to work and the ability to ask useful questions about the program operations, and
- Methodological Integrity in the production of the answers to the questions!

31

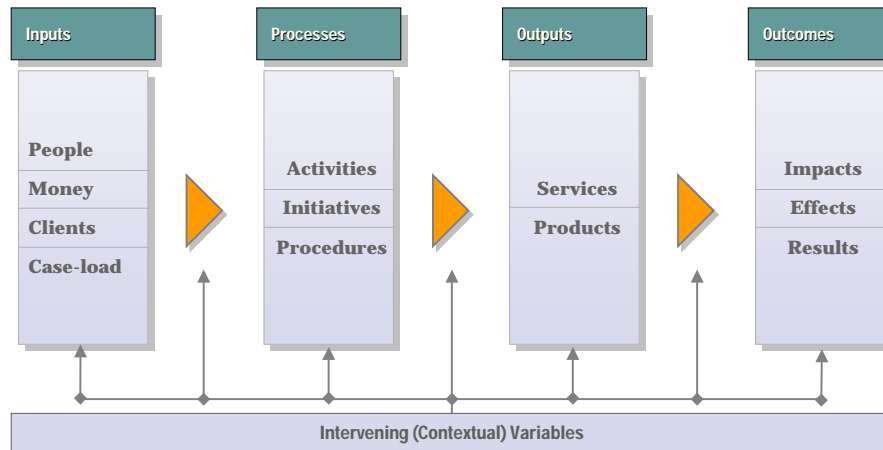
Program Logic Modeling



- Program logic models are graphic displays of the program inputs, outputs, and various outcomes that also identify key contextual variables that constrain or mediate the ability of programs to achieve desired outcomes
- These tools can facilitate communication among stakeholders, and especially between evaluators and program staff
- The tools can be used to identify where feedback loops are needed, i.e., a pathway whereby information about the results of a program is sent back to inform decisions about program resources and operations

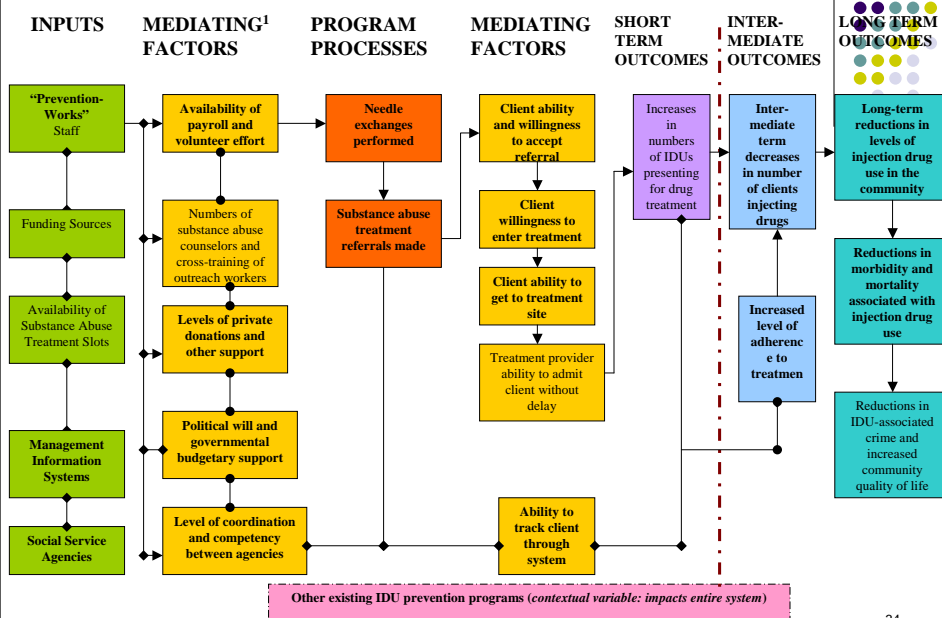
32

Generic Program Logic Model



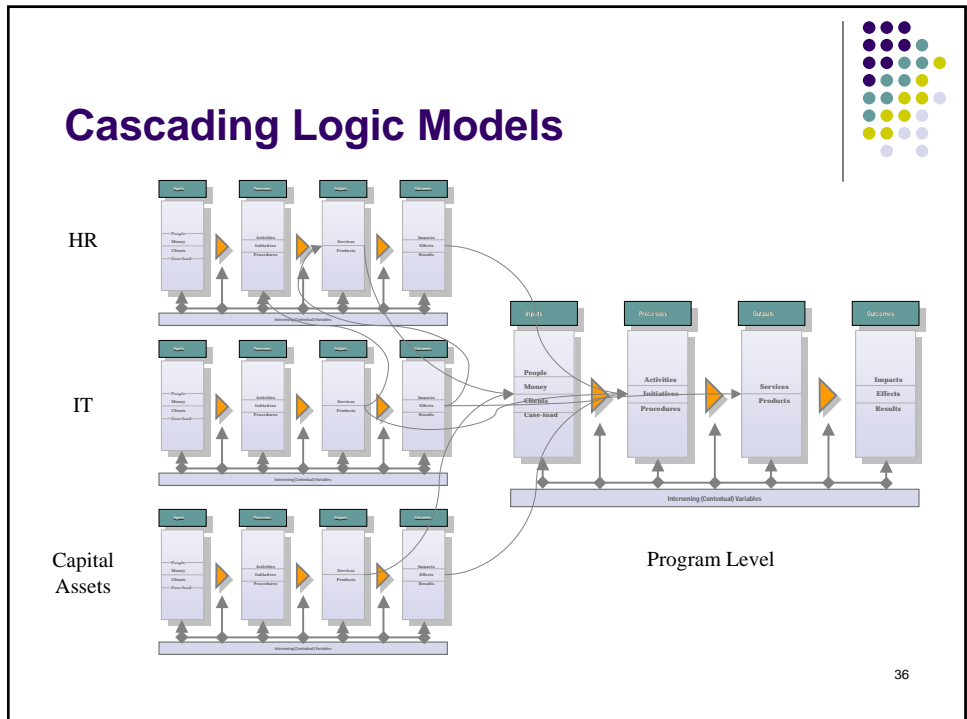
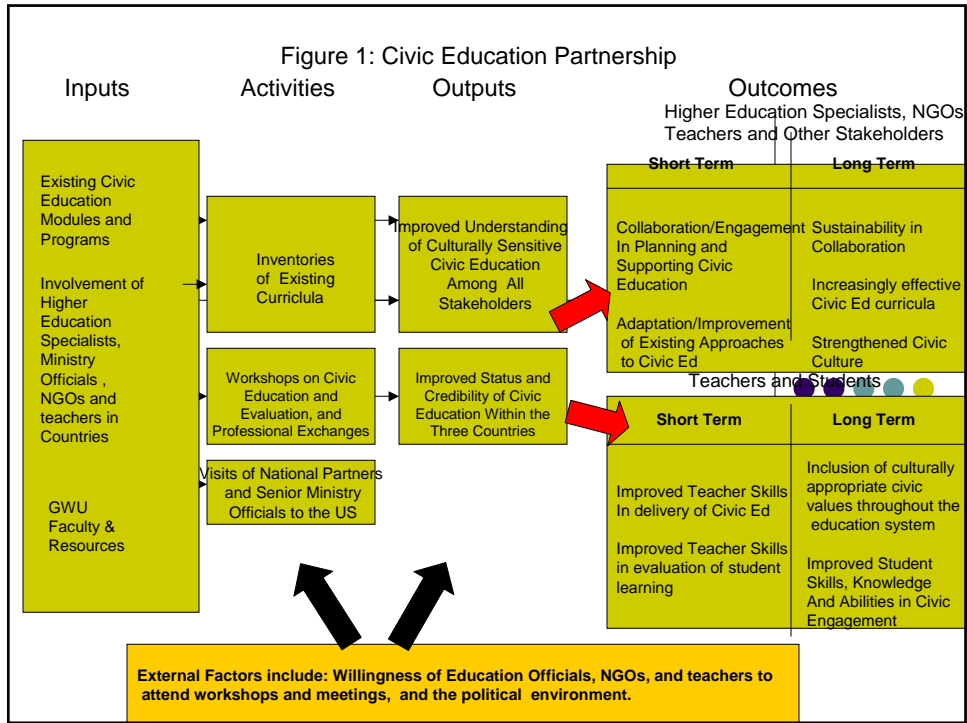
33

Program Logic: *PreventionWorks* as Gateway to Substance Abuse Treatment



34

1. "Mediating Factors" are those which, given variations in levels of their intensity and scope, will positively or negatively affect relevant processes. The red vertical dotted line indicates the point beyond which it would be impractical for *PreventionWorks* to conduct an effective and affordable evaluation





Exercise #1

- In a group of 4 participants, develop a basic logic model for a program scenario of your choice, or one provided by the instructor.

37



And then, developing strategies for making measurement and reporting decisions?

38

Measurement Considerations



- Envisioning what will be valid, reliable and credible performance measures!

39

The Rules of Evidence*



- **Competence:** Was the methodology used to collect the evidence competently executed by competent professionals?
- **Relevance:** Does the evidence address the question?
- **Sufficiency:** Is the evidence convincing to the customers/to a reasonable person?

* The Rules of Evidence from the GAO Yellowbook

40

What constitutes competence?

- *Measurement validity*
- *Internal validity*
- *Reliability*



41

What constitutes competence? - Continued

. . . and, if samples are used:

- *External validity*
- *Statistical conclusion validity*



42

How are relevance and sufficiency determined?



- Credibility! -- in the eyes of the appropriate audience!

43

Measurement Validity



- **Measurement validity** is concerned with the accuracy of measurement: Are we accurately measuring what we really intend to measure? The specific criteria for operationalizing concepts such as program outputs and outcomes, should be logically related to the concepts of interest.

44



External validity

- ***External validity*** is concerned with our ability to generalize beyond the groups or context being studied: Are we able to generalize from the results?

45



Internal Validity

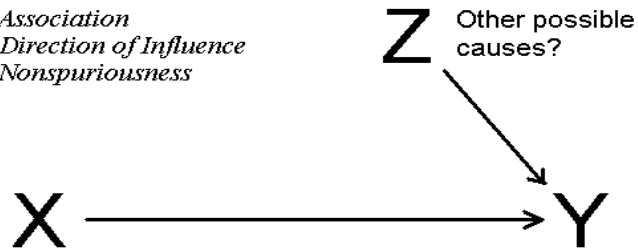
- ***Internal validity*** is concerned with our ability to determine whether X caused Y and in what magnitude: Are we able to definitely establish whether there is a causal relationship between a specific cause, such as a program, and the intended effect?

46



Causal Inference (Internal Validity)

- 1 - Association
- 2 - Direction of Influence
- 3 - Nonspuriousness



3 Elements of Causal Inference

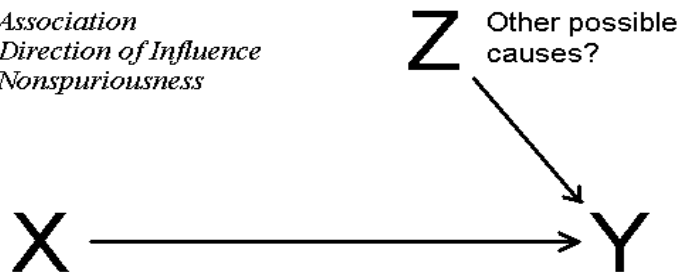
1. Temporal order
2. Co-Variation
3. Nonspuriousness

47



Programmatic Causation

- 1 - Association
- 2 - Direction of Influence
- 3 - Nonspuriousness



The Causal order presumably goes from the inputs and activities to outputs, and from outputs to short-term, then intermediate-term, and finally longer-term outcomes. However, the contextual factors that are outside the control of the program staff are the Z factors⁴⁸.

Causal Inference with Program Outcomes



- Causal inference --- unlikely
- Plausible Attribution --- maybe
- Contributions --- likely

49

Statistical Conclusion Validity



- **Statistical conclusion validity** is concerned with our ability to detect an effect, a relationship, or a factor, if it is present, and/or the magnitude of an effect with the analytical technique we employ: Do the numbers we generate accurately detect the presence of a factor, a relationship, or effect of a specific or reasonable magnitude? Numerical figures are valid if they are generated with appropriate statistical techniques supported by reasonable assumptions.

50



Reliability

- **Reliability** is the extent to which measurement procedures can be expected to produce consistent results: Will the measurement procedures produce similar results on repeated observations of the same condition or event?
- **Reliable measures** mean that operations consistently measure the same phenomena.
- **Reliable measurement** means consistently recording data with the same decision criteria.

51



Cross-Cultural Data Issues

- Collecting data cross-culturally (even within the same American city!) raises interesting issues regarding both what and how questions are framed and who does the asking!
- Cultural Competence is an issue in collection, interpretation, and reporting!

52

Credibility



- Are the performance data believable and legitimate to the intended audience?
 - Performance data are more likely to be accepted if the program stakeholders perceive the collection process is clear and unbiased.

53

Where to start: Answer Three Key Questions



- What did you do?
- How well did you do it?
- What difference did it make?

54

What Did You Do? Evaluating Implementation



Examine Activities and Outputs

- Did you conduct activities as planned?
- Did those activities produce the outputs you envisioned?

55

How Well Did You Do It? Evaluating Implementation



- What information will help you understand your program implementation?
- Think about:
 - Participation
 - Quality
 - Satisfaction

56

What difference did it make? Evaluating Outcomes



- **Outcomes:** the changes you expect to see as a result of your work
- **Indicators:** the specific, measurable characteristics or changes that represent achievement of an outcome. They answer the question: How will I know it?

57

Evaluating Outcomes: Common Types of Change



- **New knowledge**
- **Increased skills**
- **Changed attitudes, opinions, or values**
- **Changed motivation or aspirations**
- **Modified behavior**
- **Changed decisions**
- **Changed policies**
- **Changed conditions**

58

Chain of Outcomes



- **Short-term** – What changes do you expect to see?
- **Intermediate** - What changes would you want to see after that?
- **Long-term** – What changes would you hope to see after that?

59

Sample Chain of Outcomes



InnoNet's evaluation training series' outcomes:

- Expect increased knowledge of evaluation concepts (short-term)
- Want participants to apply concepts in their work; create evaluation plans and conduct evaluations (intermediate)
- Hope this will lead to program improvements for participants and their clients/audiences (long-term)

60

Sample Chain of Outcomes



CDC's evaluation training series' outcomes:

- Expect increased knowledge of evaluation concepts (short-term)
- Want participants to apply concepts in their work; create evaluation plans and conduct evaluations (intermediate)
- Hope this will lead to program improvements for participants and their clients/audiences (long-term)

61

Evaluating Outcomes: Indicators = What to Measure



Characteristics:

- Meaningful
- Direct
- Useful
- Practical

62

Evaluating Outcomes: Multiple Indicators



How many indicators are necessary?

- Simple/narrow outcomes may only need one indicator
- Complex/broad outcomes may require several

63

Evaluating Outcomes: Targets?



- Benefits of setting targets:
 - Demonstrating “stretch’ goals” to funders
 - Motivating staff
- Risks involved with setting targets:
 - Setting unrealistic targets based on hunches
 - Converting achievement into a pass/fail scenario

64

Evaluating Outcomes: Indicator Statements



Elements of a Strong Statement:

- How much
- Who
- What
- When

Outcome: Participants institute improved money management practices

Indicator: 50 of program participants open a free bank checking account within six months of beginning the program

65

Sample indicators



Outcomes	Indicators
New mothers increase knowledge of child development	<ul style="list-style-type: none"> ● 75% of participants satisfactorily complete a short test on child development at end of course
Neighborhood is a safer place for children	<ul style="list-style-type: none"> ● Violent crime decreases by 15% after one year ● 60% of residents report in one year that they believe the neighborhood is safer for children than it was before
Program resources are diversified	<ul style="list-style-type: none"> ● In 1 year, each funding source (gov't, corporations, foundations, individual donors) will comprise at least 15% and not more than 55% of program income
Increased cultural competency of legal aid attorneys	<ul style="list-style-type: none"> ● 80% of attorneys report learning about cultural issues at end of workshop ● 60% of participants' clients report that the attorneys were knowledgeable about cultural issues

66



Exercise #2

- The board of a nonprofit social service agency in DC wants to evaluate how well they are providing services to homeless youth. The services include a needle exchange program, a soup kitchen, and temporary shelter. Please identify (and justify your decisions):
 - at least three specific program outcomes you want to measure;
 - an indicator (s) for each outcome, and
 - the likely data collection method used.

67



Reporting on Use of Funds to Funders is a Real Driver in the Current Environment

- Style over Substance?
- Usefulness? For whom?

68

Sample: Outcome Funding



Results to be Achieved through the **Student Health Services** Investments:

Investments in Health Services at regular high schools and high-need middle schools will contribute toward targets for **Access to Health Care, with a Focus on Managing Health Disparities**.

Health Targets	2006	2007	2008	2009	2010	2011	2012
Citywide access to care: High school and middle school students receiving primary care	5,000	5,000	5,000	5,000	5,000	5,000	5,000
School targets in childhood immunization, focusing on south Seattle neighborhoods*	2,500	2,500	2,500	2,500	2,500	2,500	2,500
School targets in managing asthma and other chronic conditions**	600	600	600	600	600	600	600

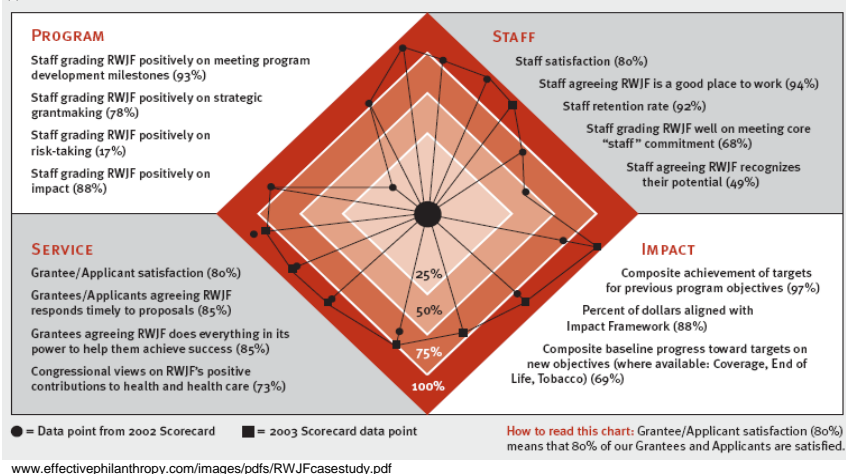
* Estimate based on School District data on students not in compliance with required immunizations as of October 2004.

** Estimate based on 2001 National Health Interview Survey data, which found that 13 percent of children under 18 years had been diagnosed with asthma. Asthma is one of the most common chronic conditions among children and has a significant impact on school absenteeism. Rates are highest among African American and low-income children.

seattle.gov/neighborhoods/education/implementation&evaluation.pdf

69

Sample: Balanced Scorecard



70

Visual: Performance Dashboard

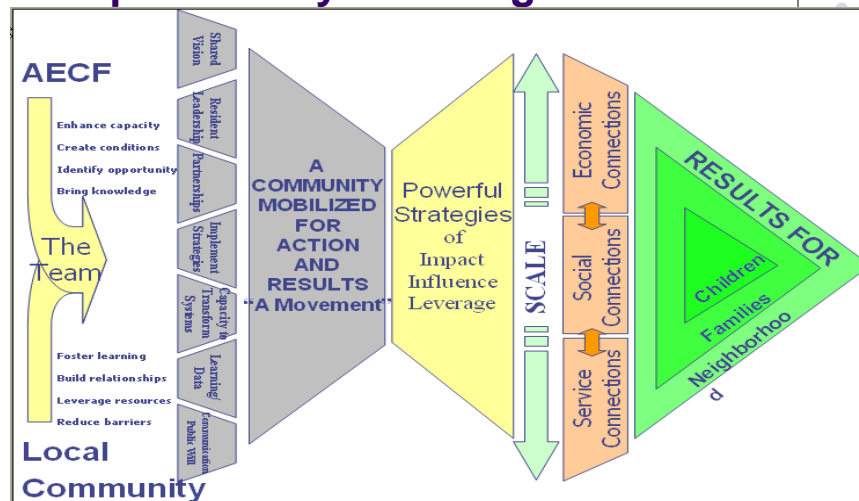


Foundation Performance Dashboard Building Blocks

Grantmaking	Programs	Operations	Investments	Social Impact
Quarterly -Discretionary grants -Board-approved grants -Requests received vs. funded -Giving vs. strategic and program targets	Quarterly -Key program developments Annual -Communications metrics	Quarterly -Payout Annual -Expense ratios -Grantee perception data -Average grant size -Communications metrics	Quarterly -Investment performance	Annual -Major initiative evaluation summaries -Progress on key value creation dimensions: grantee selection, signaling other funders, improving grantee performance, and advancing the field

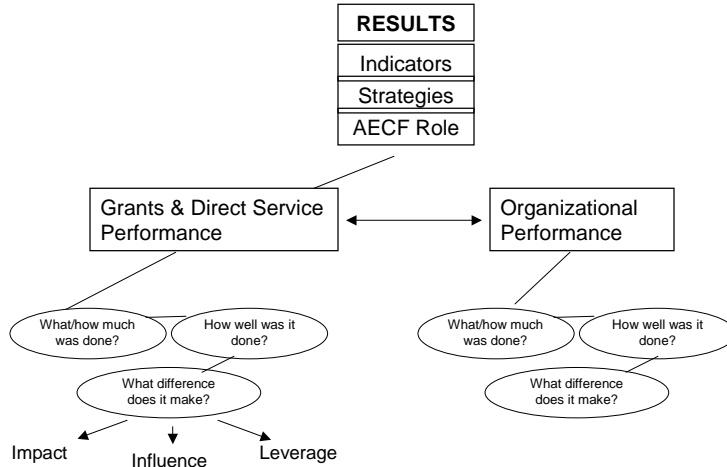
www.strategicphilanthropy.com/documents/Winter2003-private.pdf

Sample: Theory of Change



http://www.aecf.org/initiatives/mc/lp/cohort_eval/MCTheoryAug03.ppt

Foundation Approaches to Evaluation: AECF and Results-Based Accountability



73

Foundation Approaches to Evaluation: RWJF's Impact Framework



Portfolio Overview

The Foundation focuses the majority of its grantmaking in four Portfolios. For a project to be eligible for support, it must successfully address the objectives of one of the following four Portfolios.

PORTFOLIO	OBJECTIVE
Targeted	Addressing specific improvements in targeted health and health care challenges over a defined time period.
Human Capital	Attracting, developing and retaining high-quality leadership and workforce to improve health and health care.
Vulnerable Populations	Supporting promising new ideas that address health and health care problems that intersect with social factors—housing, poverty and inadequate education—and affect society's most vulnerable people, including low-income children and their families, frail older adults, adults with disabilities, the homeless, those with HIV/AIDS, and those with severe mental illness.
Pioneer	Promoting fundamental breakthroughs in health and health care through innovative projects.

<http://rwjf.org/portfolios/overview.jhtml>

74

Key Lingerin Issues



- Benefits to the Opportunity Costs of measurement?
- Meaningful findings?
- Who will fund evaluation efforts?
- Targeting, or not?
- Who will use the data for what?
- How might evaluation be useful to management as a learning tool?
- Contribution versus plausible attribution versus causation?