It is important to determine what you want to evaluate in terms of steps in program development. The next 2 slides look at structure, process [inputs and outputs], and outcomes as measures of process. Health education often suffers in evaluation programs because of the limited resources allocated to it. Corporations spend millions to increase their ‘brand share’ by 1 or 2%, however, if we increase the percentage of sexually-active individuals who use condoms by 5%, our program is often considered a failure. Unrealistic expectations can often doom a program.

There is agreement that evaluation should be comprehensive and that ideally, every aspect of a program should be evaluated. This diagram indicates some of the aspects of a program that should be evaluated and details three levels to include in the evaluation.
1. What are the key questions you would like to answer in an evaluation?

2. What is the general goal(s) of your program? (e.g., children will learn to read)

3. What is your target population? How many participants will you serve? (e.g., 40 children ages 3-4)

4. In what areas will they improve? (e.g., they will learn pre-literacy skills)

5. How will they improve? By what methods? How much will they improve? (if you have specific performance targets you would like to achieve, write them here; e.g., 40 children ages 3-4 will show 1 age-level increase in pre-literacy skills.)
6. Who will be involved in developing the evaluation?

7. How will you obtain involvement/input from key staff who will approve, administrate, and participate in the evaluation?

8. What is the group’s experience with evaluation? Can you bring in additional assistance if needed?

9. Who are your stakeholders? How will they be involved or provide input?

10. What are you, your organization, and other stakeholders going to want to learn from the evaluation?

11. What are the conditions that you are targeting with your program (what needs, problems, service gaps, etc. are you proposing to meet)?

12. If your program worked to perfection, how would these conditions change (i.e., what is your vision of the corrected or improved situation)?

13. For each of the program objectives, what are the intended immediate and or short-term outcomes for your target population (i.e., what will participants gain specifically from participating in your program? This will usually be something concrete and identifiable, such as knowledge, skills, or attitude improvement in a specific area of health (for parent education, for example), or services such as dental treatment or immunization.

---

**Session Contents**

- Basic Concepts in Evaluation
- An Evaluation Model
- Steps in Planning an Evaluation
- Stakeholders
- Evaluation Approaches
- Monitoring
- The Log Frame & PMP
- Evaluation & Experimental Design
- Testing Procedures – Validity & Reliability
- Sample Size
- Other Evaluation Links
-raithwaite et al - The development, design, testing, refinement, and application of an evaluation framework for community practice and social-professional networks
- Appendices
1. Basic Concepts in Evaluation

The slides in the beginning of this section have been taken from downloadable materials listed at the end of the session.

This section is meant as an introduction to those who have had limited evaluation experience.

We will not discuss those slides with the big white x in the corner in class.

This session also has a large Appendix and that will not be discussed in class.

What is Evaluation?

Evaluation simply means to place a value on something so it is a rather subjective activity. Evaluation is the systematic collection and analysis of data needed to make decisions, a process in which most well-run programs engage from the outset.

Here are just some of the evaluation activities that are already likely to be incorporated into many programs or that can be added easily:

1. Pinpointing the services needed For example, finding out what knowledge, skills, attitudes, or behaviors a program should address.
2. Establishing program objectives and deciding the particular evidence (such as the specific knowledge, attitudes, or behavior) that will demonstrate that the objectives have been met.
3. A set of clear, measurable, and realistic program objectives. If objectives are unrealistically optimistic or are not measurable, the program may not be able to demonstrate that it has been successful even if it has done a good job.

http://www.ed.gov/offices/OUS/PES/primer1.html
What is Evaluation?

- Developing or selecting from among alternative program approaches. For example, trying different curricula or policies and determining which ones best achieve the goals.
- Tracking program objectives. For example, setting up a system that shows who gets services, how much service is delivered, how participants rate the services they receive, and which approaches are most readily adopted by staff, or
- Trying out and assessing new program designs. For example, determining the extent to which a particular approach is being implemented faithfully by school or agency personnel or the extent to which it attracts or retains participants.

http://www.ed.gov/offices/OUS/PES/primer1.html

Types of Programs Benefitting from Evaluation

The kind of program evaluation most appropriate for a given situation depends on The Setting of the Program and The Intended Audience for the Study:

Social experiments are programs conducted as field research projects meant to test one or more hypotheses. Although they often provide services, the services are secondary in importance to the data sought by the investigators. Outcome and impact evaluations using robust quantitative measures are generally appropriate to this type of project.

Demonstration projects are not justified primarily by their research value, as are social experiments. Demonstration projects are more commonly treated as prototype programs, and process evaluation is usually conducted to determine how well the program can be implemented.

Developmental programs are usually smaller scale trials to test an innovative idea. The evaluator often wants to know if the idea can work in practice, and under what circumstances. Process evaluation is often most important in this setting as well.

New programs are settings in which new services are being offered in addition to other existing services. In this setting, the new program may not have been fully thought out at the onset and unforeseen factors may influence its implementation.

The evaluator often has little control over the program design or comes to the project too late to design an adequate study.

http://www.hcwp.org/resources/communityhealth/primer.asp

http://www.hcwp.org/resources/communityhealth/primer.asp

http://www.hcwp.org/resources/communityhealth/primer.asp
“Modification of ongoing programs” presents situations in which evaluators often have the least influence over program design and implementation. Additionally, vested interests are most prevalent in this situation, making effective evaluation challenging.

Natural experiments entail retrospective study of major changes in health programs. In this case, the program changes occur without regard to evaluation and the drawing of conclusions. In these circumstances, adequate data are seldom available.


http://www.hcwp.org/resources/commhealth/primer.asp

Types of Programs Benefitting from Evaluation

Aspects to include in a program description are:

**Need:** A statement of need describes the problem or opportunity that the program addresses and implies how the program will respond.

**Expected effects:** Descriptions of expected effects convey what the program must accomplish to be considered successful.

**Activities:** Describing program activities (i.e., what the program does to effect change) permits specific steps, strategies, or actions to be arrayed in logical sequence. This demonstrates how each program activity relates to another and clarifies the program’s hypothesized mechanism or theory of change.

**Resources:** Resources include the time, talent, technology, information, money, and other assets available to conduct program activities.

http://www.cdc.gov/eval/steps.htm#stakeholders

Aspects to include in a program description are [continued]:

**Stage of development:** Public health programs mature and change over time; therefore, a program’s stage of development reflects its maturity. A minimum of three stages of development must be recognized: planning, implementation, and effects. During planning, program activities are untested, and the goal of evaluation is to refine plans. During implementation, program activities are being field-tested and modified; the goal of evaluation is to characterize real, as opposed to ideal, program activities and to improve operations, perhaps by revising plans. During the last stage, enough time has passed for the program’s effects to emerge; the goal of evaluation is to identify and account for both intended and unintended effects.

**Context:** Descriptions of the program’s context should include the setting and environmental influences (e.g., history, geography, politics, social and economic conditions, and efforts of related or competing organizations) within which the program operates. Understanding these environmental influences is required to design a context-sensitive evaluation and will aid users in interpreting findings accurately and assessing the generalizability of the findings.

http://www.cdc.gov/eval/steps.htm#stakeholders
Aspects to include in a program description are (continued):

**Logic model**: A logic model, which describes the sequence of events for bringing about change, synthesizes the main program elements into a picture of how the program is supposed to work. Often, this model is displayed in a flow chart, map, or table to portray the sequence of steps leading to program results.

This slide and the next show pictures of what logic models are like.

This diagram lays out a typical sequencing of evaluation-linked activities.

To regulate the use of nuclear energy and establish a practical health, safety, and environmental basis to support Canada’s institutional commitments to the peaceful use of nuclear energy.
The 3 previous slides are what are called linear cause-effect models. The text on this slide and the next come from an email:

from Michael Patton <MQ Patton@PRODIGY.NET>
to: EVALSYS@LISTS.EV ALUATION.WMICH.EDU
subject: [EVALSYS] Patton on Complexity

Patton writes: “In essence, for beginners, I reduce the basic options for conceptualizing models of change to three:

1. linear cause-effect models of the kind commonly expressed in logic models where some intervention is supposed to lead, fairly directly, to some identifiable outcomes.

2. system change conceptualizations where the focus is on how sub-systems are related to each other in some system (arbitrarily and pragmatically bounded) and the focus of reforms or change is on changing the relationships and functions, and therefore, effects, of the system, and

3. nonlinear dynamics conceptualizations where some innovation is introduced into a complex situation without predetermined, predictable, or controllable results and the challenge is to track what emerges, helping those involved add resources where they see “good” things happening (based on explicit values rather than explicit predetermined outcomes) and guiding them to withdraw resources or change the innovation where they see “bad” things happening, the definitions of “good” and “bad” also being dynamic and emergent as the interactions unfold.

The first approach, I [Patton] argue, is most appropriate where a great deal is known, from research, about the likelihood of the instrumental causal connections, timeframes are reasonable to make the causal connection, those intervening have considerable control over the situation, and good measures of intervention and outcomes are available.

The second approach focuses on mapping exercises to conceptualize systems relationships, how the relationships function for certain results at baseline, and how changing relationships within the system and in the system’s relationships with other outside systems, could produce different system-level functions and results, and then mapping and tracking those changes.

The third approach has special value where knowledge about the nature of the problem is quite limited (and therefore knowledge about solutions is even more limited), where the environment is highly volatile, and/or where there’s a high value on trying things out, exploration, and innovation – that is, setting things in motion to see what happens and tracking what emerges to adapt and adjust and follow the flow.

I’ve called this “Developmental Evaluation” and am currently coaching 10 developmental evaluators in national Canadian organizations to experiment with this approach.”

Elements of an Evaluation Design

(NCI 1992) According to CDC [link below], every formal design—whether formative, process, outcome, impact, or a combination—must contain eight basic elements.

1. A Statement of Objectives
   Unless there is an adequate definition of desired achievements, evaluation cannot measure them. Evaluators need clear and definite objectives in order to measure program effects.

2. Definition of Data To Be Collected
   This is the determination of what is to be measured in relation to the objectives.

3. Methodology
   A study design is formulated to permit measurement in a valid and reliable manner.

4. Instrumentation
   Data collection instruments are designed and pretested. These instruments range from simple tally sheets for counting public inquiries to complex survey and interview forms.
Elements of an Evaluation Design [cont.]

5. Data Collection
   - The actual process of gathering information.

6. Data Processing
   - Putting the information into a usable form for analysis.

7. Data Analysis
   - The application of statistical techniques to the information to discover significant relationships.

8. Reporting
   - Compiling and recording evaluation results. These results rarely pronounce a program a complete success or failure. To some extent, all programs have good elements and bad. It is important to appreciate that lessons can be learned from both if results are properly analyzed. These lessons should be applied to altering the existing program or as a guide to planning new efforts.

http://www.atsdr.cdc.gov/HEC/evalprmr.html

Notes for Your Evaluation Design

Begin to insert notes with regard to a possible evaluation design for your program in the spaces provided in this slide and the next one:

1. A Statement of Objectives

2. Definition of Data To Be Collected

3. Methodology

4. Instrumentation

Elements of an Evaluation Design [cont.]

Begin to insert notes with regard to a possible evaluation design for your program in the spaces provided in this slide and the previous one:

5. Data Collection

6. Data Processing

7. Data Analysis

8. Reporting
Not every program should be evaluated. Following are some possible reasons.

Initially, there is need to ask whether a program can be evaluated.

Major Steps in Evaluability Assessment:


1. Involve stakeholders and intended users:
   (Note: The whole process is iterative and the stakeholders and users need to be updated on a regular basis).
   a. One user or several users? Use advisory group?

2. Clarify program intent:
   a. Interviews with key policymakers, managers, and interest group representatives?
   b. Develop a flow chart model of the program, linking resources and activities to intended immediate and longer term outcomes.
   c. Develop a list of agreed-on performance indicators [see next slide] to be used in the evaluation.
Indicators are important for evaluation.

This guy is wearing an indicator sign.

3. Explore program reality:
   Try to determine whether the program model represents reality or needs to be revised.
   a. Site visits to program.
      Interviews with operating-level managers and staff and street-level bureaucrats.
   b. Look at program documents and program data.
      Identify and verify performance measures to be used in the program.

4. Reach agreement on any needed changes in the program design.
   If evaluability assessments reveal that program objectives are implausible, given the manner in which the program currently operates, determine changes needed in model.
3. Information-based Program Objectives

5. Explore alternative evaluation designs:
   a. Outline a set of evaluation design possibilities including comparisons that could be made.
   b. Determine performance measures that could be realistically gathered.
   c. Calculate likely costs of those measurements and comparisons.
   d. Calculate costs in the time of evaluation staff, etc.

6. Agree on intended uses of evaluation information:
   a. Written Contract?
   b. Utilization plan.
REASONS FOR EVALUATING

• Collect evidence on a program’s effectiveness or impact
• Be accountable to funders, volunteers, staff, and boards
• Identify ways to improve a program by:
  1) assessing the needs of individuals, groups, communities
  2) improving the usefulness of program materials
  3) determining what works, what doesn’t, and why
  4) clarifying program plans
  5) improving communication among those involved in the program

http://www.cdc.gov/eval/evalcbph.pdf

Four standards of Program Evaluation

1. Utility (Is the evaluation useful?)
2. Feasibility (Is the evaluation viable & practical)
3. Propriety (Is the evaluation ethical?)
4. Accuracy (Is the evaluation correct?)

Guiding Questions for Program Evaluation

1. For whom is the evaluation?
2. What program are we evaluating?
3. What methods will we use in conducting our evaluation?
4. How will we gather and analyze information that is credible and in what forms?
5. How will we justify our conclusions?
6. How can we be assured that what we learn will be used?

http://www.cdc.gov/eval/evalcbph.pdf
2. An Evaluation Model

Measuring the Difference: Guide to Planning and Evaluating Health Information Outreach (September 2000)

National Library of Medicine

The following diagram can be used in determining what elements to include in the evaluation section of your business plan.

The two slides following provide examples of evaluation process and impact [summative] objectives.

Here we are evaluating not the program but the extent to which the evaluation approach is adequate.

http://nnlm.gov/evaluation/guide/
Select 2 items from the list below and indicate [in this unit’s discussion forum] how the PROCESS MEASURES are relevant to your business plan.

Examples of process evaluation objectives
- programme context and its significance
- degree of meeting the programme objectives
- analysis of the cost-effectiveness (could also be classified as impact evaluation)
- identification of key actors (stakeholders) and their roles
- strengths and weaknesses in programme design
- strengths and weaknesses in programme implementation (e.g., programme management, coordination, and staff)
- identification of barriers to successful penetration of the programme,
- exploitation of programme results and outcomes
- evaluation of low-response (non-participation)

Select 2 items from the list below and indicate [in this unit’s discussion forum] how the IMPACT MEASURES are relevant to your business plan.

Examples of impact evaluation
- changes in routine behaviour
- changes in investment behaviour
- energy savings or changes in specific consumption and consequent emission reduction
- benchmarking
- market transformation
- calculation of net impact by estimating the impact of free-rider, spill-over effect and multiplier effect
- analysis of the cost-effectiveness (could also be classified as process evaluation)

An Evaluation Model

Models are useful tools for increasing the probability that all important factors are being considered when looking at a problem.

Additionally, models suggest certain patterns of interrelations among model components.

The next few slides present a model for these purposes.
The next slide is a model for Public Health Evaluation applied to cancer and developed by Ron Anderson (now at UCLA) and LuAnn Aday (Michigan) and adapted for Health Promotion by Larry Green [at one point – Director of Extramural Programs and Academic Linkages, Centers for Disease Control and Prevention, Public Health Practice Program Office, Atlanta, GA].

The model is set up so that Assessment and Evaluation steps are from right to left and Planning and Monitoring are from left to right.

The next few slides after the model will ask you to apply the model to your organizational program.

---

**Application of the Model**

Apply the model **right to left** in the previous slide to your program:

1) How does your program aim to improve the Quality of Life of program recipients?

2) What Health/disease/illness/disability Measures will you use to determine if your program is working?
Application of the Model [cont.]

3) What are the Epidemiologic and Risk Factors your program will address?

4) What do program recipients have to do [Actions They Have to Take] to improve their health/reduce disease risk?

Application of the Model [cont.]

5) What resources have to be in place [see Enabling Factors] for the program to be implementable?

6) What basic knowledge, beliefs, attitudes, values, and commitment does the target population have [Predisposing Factors] to have to take appropriate action?

Application of the Model [cont.]

7) What types of supports [Reinforcing Factors] have to be available to prod the target population to action?

8) What types of Person Power & Training are needed for the program to be implemented [see the slide on the page after next]
9) What types of fiscal/budgetary categories will be required (see the next slide)?

10) How does this program fit into the larger picture of your organization and how is it a step in the 'Continuity of Care' process?
Steps in Program Evaluation

The steps are as follows:

• Engage stakeholders
• Describe the program
• Focus the evaluation design
• Gather credible evidence
• Justify conclusions
• Ensure use and share lessons learned

http://www.cdc.gov/eval/steps.htm

6 Steps in Planning an Evaluation

1) Identify stakeholders
2) Arrange meetings to determine who wants evaluation, what type evaluation, why and when it is desired, what resources are available
3) Assess the Evaluability of the program
4) Examine the literature
5) Determine 1) the methodology, 2) population & sampling, 3) evaluation design, 4) selection of criteria, 5) data collection approaches, 6) choice of statistics, & 7) reporting of findings
6) Present a written proposal to evaluate

This next slide represents typical steps in an evaluation and the fact that evaluation is a never ending process with feedback from one iteration directing the next.
Steps in Program Evaluation

1. Engage Stakeholders
2. Describe the Program
3. Focus on the Evaluation Design
4. Gather and Analyze Evidence
5. Justify Conclusions
6. Ensure Use and Share Lessons Learned

Why Should We Evaluate?

The next slide provides a list of reasons for evaluation.

Go through the list and rank from ‘1’ [highest] to ‘5’ the top reasons in terms of relevance for your program. In the unit’s discussion forum, explain why you ranked the items the way you did.

1) 
2) 
3) 
4) 
5) 

Evaluation can help you:

1. Document the impact of your program on your target audience
2. Focus staff on efforts to achieve common goals and shared purposes
3. Identify and share effective practices within your program
4. Communicate program results to stakeholders
5. Successfully compete for resources/funding
6. Identify staff training needs

http://www.first5santabarbaracounty.org/docs/evaluationguide.pdf
Pick 2 or 3 from the list that you feel might be the most relevant for you and indicate why and post them in the unit’s discussion forum.

1.
2.
3.

Ask the following questions:

• Who benefited from the program and how much?

• Did participants gain the intended short-term benefits of new knowledge, skills, and attitudes?

• Did participants gain the intended long-term improvements in quality of life and well-being?

• Were new program components better (i.e., more effective, more efficient, less costly) than previous models?

Ask the following questions [cont.]:

• Were clients satisfied with the services they received?

• Have we improved in meeting the needs of our target population?

• Have staff improvements (i.e., more training, higher salaries) led to better client outcomes?
Three Levels of Evaluation

• **Project Evaluation**
  Does the grantee do what it proposed?

• **Program Evaluation**
  Is the intervention program effective?
  Does it work?

• **Portfolio Evaluation**
  Is the sum of the programs effective in achieving an overall health objectives?

4. Stakeholders

We have addressed the issue of stakeholders in the past.
Review the next two slides to see if they offer additional insight for your program.
Those involved in implementing the program

Primary users of the evaluation

Those served or affected by the program

Community members
Sponsors
Collaborators
Coalition partners
Funding officials
Administrators
Managers
Staff

Clients
Family Members
Neighborhood orgs.
Academic institutions
Elected & appointed officials
Advocacy groups
Community residents

The next slide summarizes categories of real program costs comparing three different groups of stakeholders.

In designing programs, rarely are all of the costs borne by consumers considered but these costs might be the most significant in certain instances.

<table>
<thead>
<tr>
<th>Expenditure dimension</th>
<th>Stakeholder dimension</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Institution</td>
<td>Staff</td>
</tr>
<tr>
<td>Staff costs</td>
<td>Salaries, wages, pensions etc.</td>
<td>Unpaid overtime</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depreciation</td>
<td>Lecture theatre, computing equipment</td>
<td>Skills in teaching and learning development</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expenses</td>
<td>Subsistence, registration</td>
<td>Out of hours development</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overhead</td>
<td>Software licenses</td>
<td>Additional space for PC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 3. The CNL Financial Schema*
Standards for Step 1: Engage Stakeholders

<table>
<thead>
<tr>
<th>Questions</th>
<th>Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Have you included individuals and organizations that will be affected by the evaluation?</td>
<td><strong>Utility</strong>: Ensures that the evaluation is useful and answers questions that are directly relevant to users.</td>
</tr>
<tr>
<td>• Have you considered adding new stakeholders as your program evaluation is implemented?</td>
<td><strong>Propriety</strong>: Ensures that the evaluation is an ethical one, conducted with regard for the rights and interests of those involved.</td>
</tr>
<tr>
<td>• Are your team members in the evaluation planning group trustworthy and competent?</td>
<td></td>
</tr>
<tr>
<td>• Are individuals clear about what is to be done, how, by whom, and when?</td>
<td></td>
</tr>
<tr>
<td>• Is there a written understanding</td>
<td></td>
</tr>
<tr>
<td>• Have steps been taken to assure that all stakeholders and the population served will be respected and their values honored?</td>
<td></td>
</tr>
<tr>
<td>• Have conflicts of interest been discussed?</td>
<td></td>
</tr>
</tbody>
</table>

5. Evaluation Approaches

Small World Model

This model is being presented first due to its importance in understanding the spread of SARS, HIV/AIDS, Asian Bird flu, and the like.

The diagram on the left shows how infection can spread from one social network to another across the globe and lead to outbreaks that seemingly are unrelated. Evaluation of program success may depend on being able to model the spread in your program as opposed to what the theoretical spread model might look like.
Small World Model

The model in the next slide is of CDC’s efforts to build its own small world communications system to stay on top of disease spread. The link for the diagram is from the


Three experts, including CDC Director Dr. Julie Gerberding, describe major public health issues likely to emerge in the 21st century and suggest approaches to solutions.

Global warming and pollution are the result of a small world.

My wife flew through this multi-country pollution from slash and burn agriculture the same day after a plane had crashed in Malaysia from the haze.
The next two slides look at hypothetical models for the spread and the control of infectious diseases.

The three scenarios are shown diachronically when you click on the start buttons.

For simplicity, we use a simple grid network structure for illustration. Specifically, we assume that the network consists of N nodes and each node is connected to n immediate neighbours forming a rectangular grid. The behaviour of the epidemic is described in terms of four possible states of each node:

- **S**: state of being susceptible to the disease (normal person)
- **P**: state of being infected but not infecting (incubated)
- **I**: state of being infected and infecting
- **R**: state of being removed (recovered, quarantined, dead)

**Click on the links on the next slide to see the models in action. Refer back to the 4 letter codes above.**

http://cktse.eie.polyu.edu.hk/NSR-old/SARS/

The following Flash animations (require browsers with Flash plug-in) illustrate several possible outcomes corresponding to different choices of parameter values:

- **Animation 1**: Click here to see a typical simulation of propagation dynamics
- **Animation 2**: Click here to see an outbreak leading to total infection of the population
- **Animation 3**: Click here to see a controlled epidemic

The next two slides look at the spread of SARS through maps.
Flight times and flight ranges can be important information in determining areas at risk of infection. How maps are shown can influence perceptions. The world map in the next slide shows Europe in the middle. For SARS, it should have shown North America and East Asia together.

Knowing how disease spreads and monitoring its progress can be an important evaluation tool. If you can predict what will happen without a program and show what happened with a program, this can be an indicator of the effectiveness of your program.
Evaluation means to “give value to”
Thus, evaluations tend to be subjective in nature and are directed toward meeting the needs of the group financing the evaluation
The next 3 slides provide a number of definitions of program evaluation.

Examples of Definitions of Program Evaluation [from different sources]
- the key issues in program evaluation are to attempt to determine what the effects of the program have been over a specific period of time and to determine the extent to which the specific program has caused these effects [Hatry, 1980]
- program evaluation is the application of systematic research methods to the assessment of program design, implementation, and effectiveness [Chelimsy, 1985]

Examples of Definitions of Program Evaluation [from different sources] [cont.]
Best Practices:
- Transferability: Can the results be easily and effectively applied or adapted in other regions, or at different scales?
- Innovation: Is the approach truly innovative in addressing the stated problem?
- Conservation Impact: Does the strategy have a significant impact on a key conservation need or problem?
- Durability: Will the impact or effect last over time?
- Instruction: Can the practice be taught easily in different settings?
Programme evaluation is the **Systematic Assessment** of a programme, or a part of it, to:

- firstly, to assist managers and other decision makers to assess the appropriateness, efficiency, and effectiveness of a programme,
- secondly to establish the case for new programmes, or extensions to existing programmes, and
- thirdly, to decide whether the resources for the programme should be continued, increased, reduced, or discontinued.


---

A programme evaluation is commissioned by an organisation with the aim of determining how well a programme is working, and whether it needs to be approved or abandoned [Mayne, 1994]

- Evaluations should normally have elements of being

  - analytical,
  - systematic,
  - reliable,
  - issue oriented and
  - user driven


---

**Evidence-based Evaluation**

**Outcome Indicators** (tell us if intended outcomes are being achieved)

Evidence-based decision-making relies heavily on the “evidence” produced by ongoing outcome-based evaluation.

The process begins with the development of mission-driven **Strategic Objectives** with the participation of clients, staff, volunteers, donors, funders and community partners.

For each agency strategic objective, we then identify supporting department and program-level objectives.
Examples of Definitions of Program Evaluation
[from different sources – cont.]

Evidence-based Evaluation

For each program, we identify the following:
- Inputs (resources required to achieve our objectives)
- Activities (actual programs and services)
- Outputs (the products of our activities)
- Outcomes (benefits to our clients and the communities we serve)

The result of this evaluation provides the evidence upon which we make our decisions. Results also provide data for the development of new knowledge…….


1) What effects do you feel your program should have?
2) Do you feel the design of your program needs to be evaluated to see if it using an ‘evidence-based’ or ‘best-practices’ approach?
3) What types of elements of your program allow for a systematic evaluation over time?

Based on points raised in the past 3 slides,

Even when you have carried out what you feel was a good evaluation study, there are many outside reasons why others may not be as willing to accept your findings as do you.

The next slide provides a few of the possible reasons.

To paraphrase the saying:

*They shoot evaluators don’t they?*

Evaluation should be integrated into each and every program step.

The slide after next does not do that.
Barriers to Acceptance of Evaluation

1) Expecting dramatic results
2) Stakeholder pressure
3) Evaluation seen to inhibit innovation
4) Fear program will be terminated
5) Fear info will be abused
6) Fear that hard numbers won’t paint true picture
7) Fear that evaluation will drain resources
8) Fear of losing control of program
9) Fear that there will be little impact demonstrated.

The problem with the diagram on the left is that evaluation needs to be involved with each step in the process.

How would you redesign the figure to show that?

The next slide looks at how 2 counties prepared for disasters.

6. Monitoring
Monitoring

Monitoring is found in every nook and cranny of an organization now. However, only a few years ago, companies built errors directly into the cost of production and passed them on to the consumer.

The Six Sigma Approach, a way for continuous monitoring was made popular by Jack Welch of GE.

Ed Deming, who was ignored by US manufacturers until he led a revolution in Japanese productivity, helped Japanese firms implement rigorous monitoring.

Six Sigma was touched on in an earlier session.

Six Sigma is based on the realization that being right 99% of the time is just not good enough.

Medical Care is at around 2 Sigmas while baggage handling is around 4.5

What does 99% quality mean?

- 20,000 wrong prescriptions per year
- 500 incorrect surgical operations per week
- 15,000 newborns accidentally dropped per year
- 2 long or short airplane landings per day
- 2,000 lost pieces of mail per hour
- Unsafe drinking water 1 hour per month

Is 99% good enough?

http://www.itil-itsm-world.com/sigma.htm
Stewart, Deming and others used the "Plan, Do, Check, Act" circle as a guide to problem solving. (Some refer to it as the Plan, Do, Study, Act circle). Everything seems to have multiple names!

The Approach to 6 Sigmas can be seen on the next slide...

The 6 Sigmas concept means that error levels are on the outside of the area between -3 and +3 standard deviations from the mean. This is a very rare occurrence. 2 Sigmas is represented by the area in the circle. For example, something like 30% of all prescriptions have an error in them.

This means that the medical care field is closer to the 'stone age' than where we want to be.

### Basic Concepts in Monitoring

<table>
<thead>
<tr>
<th>Year</th>
<th>Clinical malaria</th>
<th>Microscopy</th>
<th>Deaths in hospitals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992</td>
<td>256,001</td>
<td>43,019</td>
<td>353</td>
</tr>
<tr>
<td>1993</td>
<td>210,830</td>
<td>42,620</td>
<td>496</td>
</tr>
<tr>
<td>1994</td>
<td>307,943</td>
<td>55,214</td>
<td>804</td>
</tr>
<tr>
<td>1995</td>
<td>311,553</td>
<td>56,118</td>
<td>804</td>
</tr>
<tr>
<td>1996</td>
<td>307,459</td>
<td>45,372</td>
<td>637</td>
</tr>
<tr>
<td>1997</td>
<td>370,529</td>
<td>59,656</td>
<td>588</td>
</tr>
</tbody>
</table>

Simple charts such as the one above can be used to see whether increased microscopy leads to reduced deaths. What do you think? How could you find out?
Program monitoring implies taking multiple measures over time. These can include:

1. Primary clinic daily appointments and/or appointments kept
2. Immunizations given at a facility over time
3. Bed days provided at a long term facilities
4. Visits made to people who are contacts of someone with an STD.

Usually, the monitoring compares achievement with a goal or objective.

The Log Frame & PMP

The next slide is of the framework discussed earlier used by USAID to justify programs and to develop the basis for an evaluation.

The rows and columns are interdependent. Goals are stated and their attainment is based on specific stated assumptions like the example below:

We will increase the use of oral rehydration therapy [ORT] for children under five because ORT has been shown to help save lives of under fives.
The Log Frame

The logical framework is based on the following below with planning from right to left and implementation from left to right.

- **Planning**
  - Inputs
  - Outputs
  - Purpose
  - Goal

- **Implementing**

Goals are set, the Purpose can include improving farming practices, Outputs can be more extension agents, and Inputs can be trainers and training sessions.

History of the Logframe Approach to Project Design

Until the early 1970s, identifying projects was a rather haphazard process and had little of the systematic thinking that had already been incorporated in standard cost-benefit methods.

In 1969, the so-called "logframe approach" [see the next slide] was developed by Practical Concepts, Inc. at the request of the U.S. Agency for International Development.

For a number of years, until the adoption of the Performance Management Plan [http://www.usaid.gov/assistance/3616], it was used by USAID and other international agencies and national governments in the planning, implementation, and evaluation of development projects.

The next slide shows that movement from one level to the next is based on assumptions.

The slide after next shows a blank Logframe Planning Matrix that you can begin completing for your organization or program and the slide after that shows the criteria that fit in each of the boxes in the model.

The slide after that is the logframe of a program in Ghana aimed at increasing functional literacy.
The next slide is a matrix that can be used to develop the logframe. You should start on the upper left and work back and forth down the matrix.

**Table 1.1: A Logical Framework Matrix**

<table>
<thead>
<tr>
<th>Narrative Summary</th>
<th>Objectively Verifiable Indicators (OVI)</th>
<th>Means of Verification (MOV)</th>
<th>Important Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOAL</td>
<td>Measure of Goal Achievement</td>
<td>Sources of Information</td>
<td>Assumption Affecting</td>
</tr>
<tr>
<td>PURPOSE</td>
<td>End of Project Status</td>
<td>Methods Used</td>
<td>Purpose-Goal Linkage</td>
</tr>
<tr>
<td>OUTPUTS</td>
<td>Magnitudes of Outputs</td>
<td>Sources of Information</td>
<td>Assumption Affecting</td>
</tr>
<tr>
<td></td>
<td>Planned Completion Date</td>
<td>Methods Used</td>
<td>Output-Purpose Linkage</td>
</tr>
<tr>
<td>INPUTS</td>
<td>Nature and Level of Resources</td>
<td>Sources of Information</td>
<td>Assumptions Affecting</td>
</tr>
<tr>
<td></td>
<td>Necessary Cost</td>
<td>Methods Used</td>
<td>Inputs-Outputs Linkage</td>
</tr>
<tr>
<td></td>
<td>Planned Starting Date</td>
<td>Sources of Information</td>
<td>Initial Assumptions</td>
</tr>
</tbody>
</table>

**Logframe Matrix**

<table>
<thead>
<tr>
<th>Goal</th>
<th>Purposes</th>
<th>Inputs</th>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Relationship between assumptions & objectives**

ONLY IF THESE ASSUMPTIONS ARE MET WILL THE NEXT LEVEL OF OBJECTIVES BE ACHIEVED. ACHIEVEMENT IS AN IMPORTANT PART OF THE VERTICAL LOGIC.
Operationalizing Concepts

The next slide was developed from a road project we completed on the Palestinian West Bank. It is laid out very similarly to the logical framework.

There is:
- a program rationale
- the intent [program outcome]
- observations as to achievements of the program
- standards by which outcomes are assessed, and
- a statement of who will be making the judgments as to what has been achieved.

The program output in the slide is "building a road"
Fill in the spaces in the next slide for the operationalizing of your own program.

USAID Performance Management Plan

The next series of slides introduces the USAID Performance Management Plan, the successor to the LogFrame.

A brief introduction and several examples are included.

The following link will provide a more extensive overview:


Here is another example:

The next few slides look at a multisectoral program to improve maternal and child health in West and Central Africa with a focus on preventing HIV/AIDS.

The Santé Familiale et Prévention du SIDA (SFPS) project is a regional initiative funded by USAID in West and Central Africa (WCA) which is dedicated to achieving four principal results:

1. Increased use of modern family planning methods to improve the health and economic well-being of all family members;
2. Increased use of condoms to reduce the transmission of HIV/AIDS and STDs;
3. Increased use of ORS to reduce child morbidity and mortality due to diarrhoeal diseases;
4. Reinforced managerial, financial and technical capacity of regional African institutions to improve their contribution to development of West & Central Africa.

English translation of “Santé Familiale et Prévention du SIDA” — “Family Health and AIDS Prevention”

http://www.tulane.edu/~sfps/sante.htm
SFPS is achieving its anticipated results by:
- increasing demand and supply of high quality, integrated family planning services
- increasing knowledge of HIV/AIDS & STI prevention practices while increasing the availability of condoms
- increasing awareness and availability of Oral Rehydration Salts (ORS), and
- creating long-term partnerships with and between national and regional African institutions in West and Central Africa.

SFPS continues to examine and seize collaborative opportunities with other donors and partners for maximum positive impact from coordinated program planning and implementation.

Input/Process Indicators:
- Increase in # of African Consultant-weeks used by SFPS
- Increase in # of Development partners collaborating with SFPS
- Increase in % of financial resources mobilized with regard to SFPS LOP commitments
- Increase in # of studies conducted by SFPS regional/national partner institutions

The next few slides provide a few vignettes of the range of SFPS activities.

SFPS selected 8 clinical training sites in 3 of the four SFPS countries to reinforce in the following areas:
- adequate number of clients,
- adequate infrastructure,
- adequate number of competent service providers,
- commitment to being a training center,
- adequate equipment and training materials and quality services (including infection prevention), and
- clinic management.
During FY2000, SFPS inputs to these sites were considerable and included training of providers, community-based information campaigns to increase clients, provision of training materials and continued supervision to ensure quality of services.

Only 3 of the 8 have met the JHPIEGO minimum standard to date. It is most probable that 19 sites will meet the minimum standard. Consequently, the target of 20/38 will not be met because the program in Côte d’Ivoire has been on hold and a second consultation needs to be constructed in two sites in Cameroon.


REDSO/PHN provides support to the region through three African partner institutions in four areas:

1) strengthening the institutional capacity and sustainability of regional partner institutions;
2) expanding the base of human, technical and program resources available to improve systems throughout the region;
3) increasing the analysis, dissemination, and application of information to enhance sector programs;
4) enhancing the policy environment by improving regional dialogue on regional policy issues in key technical areas that include HIV/AIDS, maternal and child health and nutrition, reproductive health, infectious disease, and health care financing.

[PDF](http://pdf.usaid.gov/pdf_docs/PDACD250.pdf)
These four areas of support correspond with the four Intermediate Results (IRs) REDSO identified to achieve SO7:

IR 7.1: Improved Viability of Regional Partner Institutions
IR 7.2: Broadened Technical Resource Base
IR 7.3: Expanded Utilization of Critical Information, and
IR 7.4: Expanded Policy Dialogue.

Each of the Intermediate Results has related activities as can be seen in the figure in the next slide.


8. Evaluation & Experimental Design
This section has been moved to the Appendices.

9. Testing Procedures
   - Validity & Reliability

This section has been moved to the Appendices.
This next slide presents a very useful chart which gives you help in determining an appropriate sample size for your program [it may take a few seconds to load].

Look in the left column.

This deals primarily with mortality but it can be used for morbidity and for coverage as seen in the last entry. Much smaller sample sizes are needed for health care measures and health behaviors than for health outcomes.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>p before</th>
<th>p after</th>
<th>p ratio</th>
<th>n Needed</th>
<th>Total (%) &lt;5 needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mortality 0-9 y</td>
<td>0.01</td>
<td>0.12</td>
<td>0.06</td>
<td>2.451</td>
<td>2.457</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.13</td>
<td>0.06</td>
<td>2.012</td>
<td>2.014</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>0.14</td>
<td>0.06</td>
<td>1.675</td>
<td>1.675</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.15</td>
<td>0.06</td>
<td>1.353</td>
<td>1.353</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.16</td>
<td>0.06</td>
<td>1.034</td>
<td>1.034</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>0.17</td>
<td>0.06</td>
<td>0.727</td>
<td>0.727</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>0.18</td>
<td>0.06</td>
<td>0.454</td>
<td>0.454</td>
</tr>
<tr>
<td>2. Mortality 10-14 mos</td>
<td>0.02</td>
<td>0.10</td>
<td>0.05</td>
<td>3.583</td>
<td>3.587</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>0.11</td>
<td>0.05</td>
<td>2.831</td>
<td>2.834</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.12</td>
<td>0.05</td>
<td>2.118</td>
<td>2.118</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.13</td>
<td>0.05</td>
<td>1.454</td>
<td>1.454</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>0.14</td>
<td>0.05</td>
<td>0.854</td>
<td>0.854</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>0.15</td>
<td>0.05</td>
<td>0.503</td>
<td>0.503</td>
</tr>
<tr>
<td>3. Infant mortality (&lt;3</td>
<td>0.05</td>
<td>0.11</td>
<td>0.06</td>
<td>2.453</td>
<td>2.457</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>0.12</td>
<td>0.06</td>
<td>2.013</td>
<td>2.014</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>0.13</td>
<td>0.06</td>
<td>1.675</td>
<td>1.675</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>0.14</td>
<td>0.06</td>
<td>1.353</td>
<td>1.353</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>0.15</td>
<td>0.06</td>
<td>1.034</td>
<td>1.034</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.16</td>
<td>0.06</td>
<td>0.727</td>
<td>0.727</td>
</tr>
<tr>
<td>4. Neonatal mortality</td>
<td>0.02</td>
<td>0.05</td>
<td>0.03</td>
<td>5.985</td>
<td>5.987</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>0.06</td>
<td>0.03</td>
<td>4.871</td>
<td>4.873</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.07</td>
<td>0.03</td>
<td>3.754</td>
<td>3.757</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.08</td>
<td>0.03</td>
<td>2.641</td>
<td>2.643</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>0.09</td>
<td>0.03</td>
<td>1.534</td>
<td>1.536</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>0.10</td>
<td>0.03</td>
<td>0.443</td>
<td>0.445</td>
</tr>
<tr>
<td>5. Vaccination</td>
<td>0.06</td>
<td>0.10</td>
<td>0.06</td>
<td>3.583</td>
<td>3.587</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>0.11</td>
<td>0.06</td>
<td>2.831</td>
<td>2.834</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>0.12</td>
<td>0.06</td>
<td>2.118</td>
<td>2.118</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>0.13</td>
<td>0.06</td>
<td>1.454</td>
<td>1.454</td>
</tr>
<tr>
<td></td>
<td>0.10</td>
<td>0.14</td>
<td>0.06</td>
<td>0.854</td>
<td>0.854</td>
</tr>
<tr>
<td></td>
<td>0.11</td>
<td>0.15</td>
<td>0.06</td>
<td>0.503</td>
<td>0.503</td>
</tr>
</tbody>
</table>
These are footnotes for the previous slide

- \( S = \frac{N}{\text{Total} - N} \) where \( N \) is the number of items to be sampled.
- The proportion of items to be sampled is \( \frac{N}{\text{Total} - N} \).

The End

Appendices

A) International Code of Marketing
B) Testing Procedures – Validity & Reliability
C) Threats to Validity
D) Other Evaluation Links
The International Code For Marketing has been introduced here since evaluation is all about placing a 'Value on Something'.

The International Code focuses on collecting information on and from children, the most vulnerable members of society. At the same time, the procedures used for children should be applied to the rest of the population as well.

Evaluation is a Subjective Process with all of the negatives that that can bring.

This presentation is to help stakeholders look at program achievements with as neutral a view as is possible.

www.who.int/nutrition/publications/icode_e.pdf

International Code for Marketing

Read the International Code [click on the link at the bottom of the next slide]. Respond to the following:

1) Does your program serve minors?
2) What types of information are collected from minors?
3) What guarantees are there that the information will not be misused?

www.who.int/nutrition/publications/icode_e.pdf
In carrying out an evaluation or research:

- the welfare of the children and young people themselves is the overriding consideration – they must not be disturbed or harmed by the experience of being interviewed.

- the parents or anyone acting as the guardian of any child or young person taking part in a research project must be confident that the latter’s safety, rights and interests are being fully safeguarded.
• the interviewers and other researchers involved in the project must be protected against any misunderstandings or possible allegations of misconduct arising from their dealings with the children or young people taking part in that project.

http://www.itcilo.it/english/actrav/telearn/global/ilo/guide/iccmar.htm

• the authorities and the public generally must be confident that all research carried out with children and young people is conducted to the highest ethical standards and that there can be no question of any possible abuse of the children or young people involved.

http://www.itcilo.it/english/actrav/telearn/global/ilo/guide/iccmar.htm

B) Testing Procedures – Validity & Reliability
**Evaluation Validity**

The next 2 slides look at:

1) **Internal Validity** – the extent to which your program measurements are accurate

2) **External Validity** – the extent to which you can apply findings from your program evaluation to similar programs.

---

**Internal Validity**

Internal validity is one important type of research validity. The term "internal validity" refers to the extent that extraneous variables (error variance) in an experiment are accounted for.

It is paramount to the researcher that model specification error variance (as distinct from measurement and sampling error variance) is controlled.

*Asking people how often they use health care services in a year might have a large error factor in it. Asking them how often they were hospitalized last year probably has a small level of error.*

---

**External Validity**

This construct asks the question of generalizability. Which populations, settings, treatment variables and measurement variables can these results be generalized to?

Generalizing across persons requires research samples to be representative of the population of interest. Generalizing across times and settings usually necessitates systematically administering the experimental procedure at different times and different settings.

*For example, comparing health service usage in a rural area with an urban area or between young and old populations could lead to incorrect conclusions.*
The extent to which you have to rely on testing procedures can affect your ability to evaluate your program.

- In an ACS program in Baltimore to reduce cervical cancer, 20% of the test results from one lab were found to be inconclusive as to possible cancer.
- In a blood pressure control program, people were found to have elevated pressures due to the stress of the clinical environment.

The next slide looks at these issues [blood pressure readings tend to have either even values or values bunched at the ‘0’ and ‘5’ digits].

### Test Results Table

<table>
<thead>
<tr>
<th></th>
<th>Test Positive</th>
<th>Test Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease</td>
<td>TP</td>
<td>FN</td>
</tr>
<tr>
<td>Normal</td>
<td>FP</td>
<td>TN</td>
</tr>
</tbody>
</table>

true positive (TP), false positive (FP), true negative (TN), and false negative (FN)

This chart allows you to estimate where your results will fall. You can determine percentages by doing a web-based literature review and plugging in the percentages of TP, FP, TN, & FN you find.

Using these terms, the following metrics can be defined.

- Sensitivity = TP/(TP+FN)
- Specificity = TN/(TN+FP)
- Positive predictive value (PPV) = TP/(TP+FP)
- Negative predictive value = TN/(TN+FN).
Sensitivity - the extent to which the cases that are found are real cases

Specificity - the extent to which the negatives that are found truly are negatives

PPV - the percentage of times that a positive test result will be a true positive

NPV - the percentage of times that a negative test result is a true negative

Click on the link below for a Bayesian calculator into which you can plug the values specified in the previous slide to get
• Sensitivity
• Specificity
• Prevalence
and other items

http://psych.fullerton.edu/mbirnbaum/bayes/BayesCalc.htm

In testing, we want to maximize the true positives and true negatives

To what extent does your program have to rely on test [those that have been developed elsewhere generally] outcomes to assess its success?

What does the literature [http://scholar.google.com/] suggest about test reliability issues for your program?
Experimental Design can be quite different from Evaluation Design [it also can be a subset of evaluation design]. The chart in the next slide lists requirements:

- There must be an hypothesis that something is equal to, less than, or more than something else
- There must be a situation where there is an intervention (input) and a virtually equal situation where the intervention is lacking
- The people or items either receiving the intervention or an alternate must have been allocated to the two groups in an unbiased manner.
- Threats to internal validity must be reduced to a minimum or eliminated

Evaluation & Experimental Design

Designs for Evaluation are used in implementing programs that have been shown in the past to have value.

Designs for Experimentation are needed to test innovation approaches since they are more rigorous.

The next 2 slides are of designs often used in program evaluation.

The 3rd and 4th slides are more often used in evaluating experimental research. The 5th slide is a more involved summary of slide 4.
Pre-Experimental Designs

Pre-experimental designs are those in which there is no control group and/or have comparison groups that are formed nonrandomly, therefore yielding results which are difficult to interpret (Huck & Cormier, 1996).

The three pre-experimental designs presented by Campbell and Stanley (1963) are the one-shot case study, the one-group pretest-posttest design, and the static group comparison.

![Diagram of Pretest-posttest control group design]

**Pretest-posttest control group design**

The design is diagrammed as follows:

\[ \begin{align*}
R & \quad O_2 \quad X \quad O_1 \\
R & \quad O_3 \quad O_4
\end{align*} \]

Random assignment is employed to both groups, and both are given a pretest. One group is administered the X, and the other is not. A comparison of O₂ and O₄ should elucidate any effect of the X.

![Diagram of Posttest-only control group design]

**The posttest-only control group design**

Here, randomization is utilized to ensure the equalization of the two groups, without a pretest. This is done where it is felt that the pretest will affect the way people act or feel.

The design is depicted in this way:

\[ \begin{align*}
R & \quad X \quad O_3 \\
R & \quad O_2
\end{align*} \]

Again, random assignment is employed to both groups. The X is administered to the experimental group, and the second group acts as a control.
The Solomon four-group design

When a pretest is desired, but there is concern over the effects of using a pretest, as in the Bateman, Sakano, and Fujita (1992) study, this design is used, notationally described as:

\[\begin{align*}
R & \quad O_1 \quad X \quad O_2 \\
R & \quad O_3 \quad O_4 \\
R & \quad X \quad O_5 \\
R & \quad O_6
\end{align*}\]

In this way, the differences in scores of those randomly selected to get the pretest and post test from those who get the post test only should represent the impact of the pretest if the groups are otherwise equal.

The next slide shows a lot of the variables that go into determining the **Content of an Experiment**.

The slide after that looks at design, how to arrange the content so that its impact can be accurately assessed.

Much of what we have been talking about in an evaluation to date is content-related.

The third slide discusses several experimental design formats.
Experimental Design Factors

The design on the next slide is the most ideal since it allows for assessing the impact of data collection on outcomes.

Threats to Validity

The timing of When Data Collection Is Done can be summarized as follows:

• Pre-post compared to normative data
• Post test only – intervention and control [multiple intervention and control units]
• Pre/post test intervention and control [multiple intervention and control units]

The design on the next slide is the most ideal since it allows for assessing the impact of data collection on outcomes.

The designs in the previous slides are well represented by the flow chart at the right which provides decisions as to why to select one design over another.

The important thing to remember is that experimental designs are very expensive to use. ALWAYS use a simpler design if you can.

www.coe.sdsu.edu/.../12-experimentdesign/index.htm
The slide after next provides a useful summary sheet of design issues moving from the least to the most involved.

To what extent are the following issues you need to consider in evaluating your program?

1) **History** – the extent to which the surrounding [larger] environment will affect your outcome?

2) **Maturation** – the effects of your population’s aging in the program.

3) **Mortality** – the extent to which people in your program may drop out or be lost to follow up.

4) **Testing** – the extent to which tests you might use might have an effect on program participants [raise their awareness, for example]

5) **Selection** – the extent to which those included in your evaluation are different from people in the overall program [people who are tested compared with those who are not]

6) **Instrumentation** – the extent to which tools you use [Snellen test] might not be understood by everyone [the illiterate for example]

The next three slides look at the effects of threats to validity on each of the specific designs.

In this way, you can determine what the greatest threats might be and select a design accordingly.
### Designs and Threats to Internal Validity

#### Non-experimental Designs

<table>
<thead>
<tr>
<th>Design</th>
<th>History</th>
<th>Maturation</th>
<th>Mortality</th>
<th>Testing</th>
<th>Selection</th>
<th>Instrumentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 shot case study</td>
<td>XO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 group pretest-posttest</td>
<td>OXO</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Static group compared</td>
<td>XO O</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Quasi-experimental Designs

<table>
<thead>
<tr>
<th>Design</th>
<th>History</th>
<th>Maturation</th>
<th>Mortality</th>
<th>Testing</th>
<th>Selection</th>
<th>Instrumentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonequivalent control group</td>
<td>OXO</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Separate sample pretest-posttest</td>
<td>OXO</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

[R] = random assignment

#### True Experimental Designs

<table>
<thead>
<tr>
<th>Design</th>
<th>History</th>
<th>Maturation</th>
<th>Mortality</th>
<th>Testing</th>
<th>Selection</th>
<th>Instrumentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest-posttest control group [R]</td>
<td>OXO</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Pretest only Control Group [R]</td>
<td>XO O</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

[R] = random assignment
## Factors affecting goals & objectives

<table>
<thead>
<tr>
<th>Factors</th>
<th>Your Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extent of political/moral feasibility</td>
<td></td>
</tr>
<tr>
<td>Timeliness (is your program a part of a trend or are you bucking the trend?)</td>
<td></td>
</tr>
<tr>
<td>Cost (what evidence is there that your organization can afford the program?)</td>
<td></td>
</tr>
</tbody>
</table>

Here are some questions to ask before finalizing your evaluation design. Please ask questions.

- Why might you have two ‘post test only’ groups?

**Sample size** is a function of incidence/prevalence rates

Need to consider internal and external validity and reliability in selecting measures.

An important issue in your objectives is to determine the frequency of collecting and reporting data.

---

### C) Threats to Validity
Threats to internal validity

This slide and the following 4 slides go over factors that can be a part of the experimental design that will indicate incorrect outcomes. These few slides will not give you an understanding but just an awareness of some of the issues to consider.

There are many different ways that the internal validity of a study can be threatened or jeopardized. A list and brief comment of some of the more important ones are given below.

1) Selection bias.
Occurs when more of one type of person gets into one group for a study. For example, the people who return your questionnaire may be different, in some important way, to the people who did not return your questionnaire. The students who volunteer for your project might be different to the ones who do not volunteer (for example, more altruistic, more achievement oriented, more intelligent). Do these variables have an effect on the thing you are trying to measure? We usually do not know.

2) Drop-out.
More of one type of person may drop out of one of the groups. For example, those less committed, less achievement-oriented, less intelligent.

3) History.
Events that happen to participants during the research which affect results but are not linked to the IV. In an extended study comparing relaxation to no relaxation on headache occurrence, those in the no relaxation condition sought out other means of reducing their headache occurrence (e.g. took more pills).
4) Reliability of measures and procedures.

Unreliable operationalisations of constructs, or inconsistency in giving instructions to participants, or training to assessors can invalidate the study.

5) Using a design of low power.

In particular, a small sample size may have insufficient power to detect a real effect even if it is there. As a result, the researcher claims the manipulation had no effect when in fact it does; he just couldn't pick it up. As well, different statistical tests have varying sensitivity to detect differences.

http://www.une.edu.au/WebStat/unit_materials/c2_research_design/validity_threat_internal.htm

6) Order effects.

If we measure something over a series of trials, we might find that a change occurs because our participants are becoming bored, tired, disinterested, fatigued, less motivated than they were at the beginning of the series. “Counterbalancing” is a way of overcoming this problem in repeated measures designs.

7) Multiple tests of significance.

The more significance tests (Chapter 6) you conduct on the one set of data, the more likely you are to claim that you made a significant finding when you should not have. You will be capitalising on chance fluctuations.

http://www.une.edu.au/WebStat/unit_materials/c2_research_design/validity_threat_internal.htm

D) Other Evaluation Links
## Evaluation materials online

The materials in the following slide can be viewed by clicking on the link.

### The following handbooks and texts are available in their entirety online. Most are multi-chapter documents focusing on "how to's" of evaluation-related subjects.

<table>
<thead>
<tr>
<th>Title</th>
<th>Sponsor</th>
<th>Scope</th>
<th>Audience</th>
<th>Format</th>
<th>Link</th>
</tr>
</thead>
</table>
Title: Framework for Program Evaluation  
Sponsor: Centers for Disease Control and Prevention  
Scope: Steps to Evaluation and Standards for Effective Evaluation in Public Health  
Audience: Public Health Professionals  
Format: Webpages  
Link: http://www.cdc.gov/eval/framework/index.htm

Title: HyperStat Statistics Textbook  
Sponsor: Rice Virtual Lab  
Scope: Comprehensive, interactive, novice-advanced level statistics textbook  
Audience: Program managers  
Format: Webpages  
Link: http://davidmlane.com/hyperstat/

Title: An Evaluation Primer on Health Risk Communication Programs and Outcomes  
Sponsor: US Department of Health and Human Services Environmental Health Policy Committee  
Scope: Focuses on planning for and executing evaluations of health risk communication programs and materials  
Audience: Risk communication practitioners and decision-makers in Public Health Service and non-Public Health Service agencies  
Format: Webpages  
Link: http://www.atsdr.cdc.gov/HEC/evalprimer.html

Title: Locating the Energy for Change: An Introduction to Appreciative Inquiry  
Sponsor/Authors: Charles Elliott, Cambridge College & International Institute for Sustainable Development  
Scope: Introduction to appreciative inquiry with case studies from development settings  
Audience: Development Practitioners  
Format: PDF  
Note: Chapter 12 focuses on using Appreciative Inquiry in evaluation

Title: Measuring the Difference: Guide to Planning and Evaluating Health Information Outreach  
Sponsor/Authors: Catherine Burroughs, National Library of Medicine  
Scope: Overall guide to evaluating health outreach programs  
Audience: Health Educators and Health Science Librarians  
Format: Webpage linking to PDFs  
Link: http://nnlm.gov/evaluation/guide/

Title: Qualitative Methods Workbook  
Sponsor/Author: C George Boeree, Shippensburg University  
Scope: Broad look at qualitative methods including phenomenology, structural analysis, observation and interviewing  
Audience: Students/Faculty in introductory qualitative methods courses  
Format: Webpages  
Link: http://www.ship.edu/~cgboeree/qualmeth.html

Title: Questionnaire Design and Analysis Workbook  
Sponsor/Author: Alison Galloway, Queen Margaret College  
Scope: Hands-on with activities approach to learning questionnaire design  
Audience: Broadly applicable but originally written for online course  
Format: Webpages  
Link: http://www.tardis.ed.ac.uk/~kate/qmcweb/qcont.htm

Title: Practical Guide to Monitoring and Evaluation of Rural Development Projects  
Sponsor/Authors: International Fund for Agricultural Development (IFAD) Office of Evaluation and Studies  
Scope: Development, installation and use of effective project-level M&E systems as tools for impact-oriented management  
Audience: Directors and M&E Offices of IFAD supported projects  
Format: Webpage  
Link: http://www.ifad.org/evaluation/e/process/guide/index.htm
Title: Results Oriented Monitoring and Evaluation
Sponsor: United Nations Development Programme (UNDP)
Scope: Comprehensive guide to Monitoring and Evaluation of Development Programs
Audience: UNDP Program Managers
Format: Webpages

Title: Sampling Workbook
Sponsor/Author: Alison Galloway, Queen Margaret College
Scope: Hands-on with activities approach to learning sampling
Audience: Broadly applicable but originally written for online course
Format: Webpages
Link: http://www.tardis.ed.ac.uk/~kate/qmcweb/scont.htm

Title: Taking Stock: A Practical Guide to Evaluating Your Own Programs
Sponsor: American Association for the Advancement of Science
Scope: How to conduct participatory, internal evaluations in community-based organizations
Audience: Staff in community-based organizations
Format: Webpages

DAC work on development evaluation started in 1982 with the creation of an Expert Group which was then replaced by the DAC Working Party on Aid Evaluation. Since March 2003, this work is carried out by the DAC Network on Development Evaluation. The Network provides a forum where evaluation experts and managers from 30 bilateral and multilateral development agencies meet regularly to improve the practice and use of evaluation in development assistance.

http://www.oecd.org/about/0,2337,en_2649_34435_1_1_1_1_1,00.html
OE evaluates IFAD’s projects and programmes to assess what works and what doesn’t and to determine how far IFAD’s policies and strategies are successful in tackling poverty alleviation in rural areas. In identifying key insights and recommendations drawn from evaluation findings, OE is also concerned, in accordance with IFAD’s disclosure policy, with communicating and sharing IFAD’s knowledge and experience of rural and agricultural development with a wider audience. OE’s evaluation is based on a coherent set of evaluation methodologies.

The Office of Evaluation (OE) is responsible for evaluating IFAD’s operations and policies. In April 2003, the Executive Board approved IFAD’s new Evaluation Policy, which led to OE’s new independent status: OE now reports directly to the Executive Board.

Evaluation Links

The link at the bottom of this slide will guide you to a number of websites of organizations focusing on evaluation.

If you have interest in becoming an evaluator, it may be useful to visit 1 or more sites since they represent organizations that are very much like ‘craftsmen’s guilds’.

12. Braithwaite et al - The development, design, testing, refinement, simulation and application of an evaluation framework for communities of practice and social-professional networks
Framework and Designs for Evaluation

Figure 3: Conceptual framework, design, and methods