



# **A Field Guide to Designing Quantitative Instruments to Measure Program Impact**

Dr. Susan J. Barkman  
Professor, Department of 4-H Youth  
Purdue University  
AGAD Building, Room 229  
615 W. State Street  
W. Lafayette, IN 47907-2053  
765/494-8436  
sbarkman@purdue.edu

**Acknowledgements:** “A Field Guide to Designing Quantitative Instruments to Measure Program Impact” was written by Susan J. Barkman, Professor, Purdue University. A special thanks to Krisanna Machtmes, Assistant Professor, Louisiana State University, for minor contributions. Page layout template was designed by Russ Merzdorf, Purdue University Agricultural Communications. Editing was done Frank Koontz, Purdue University Agricultural Communications.

Copyright (2002) by Purdue Extension, West Lafayette, Indiana 47907. All rights reserved. Unless permission granted by author, this material shall not be copied, reproduced or coded for reproduction by any electrical, mechanical or chemical processes, or combinations thereof, now known or later developed.

## **PURPOSE OF THIS FIELD GUIDE**

This field guide is designed to help staff in community-based organizations *with little or no evaluation expertise* develop quantitative evaluation instruments to help them assess the impact of their educational programs.

The guide ties into the Logic Model process and focuses on *outcome-based evaluation* – whether the program achieved its targeted outcomes. It does deal with unintended outcomes that may occur as a result of a program. The guide describes easy-to-follow steps for developing tests, surveys/questionnaires, and observation checklists. It includes tips on such things as writing effective questions, choosing the appropriate response choices and/or scaling, and selecting methodology for administering the instrument. Also included are quick "field" methods to help improve instruments and increase the reliability and validity of the data gathered.

**SPECIAL NOTE:** It is recommended that professionals and researchers with more experience in evaluation methodology utilize statistical procedures to establish the psychometrics (reliability and validity) for the instruments they are designing. The psychometric tests to run are briefly outlined under Step 6: Pilot Test and Revise the Instrument.

### **Steps in Designing an Outcome-Based Quantitative Instrument to Measure Program Impact**

---

- 1. Clearly identify the outcome(s) you want to achieve.**
- 2. Determine how you will measure the achievement of that outcome(s) - indicators and data sources.**
- 3. Determine the specific details you want to measure under each outcome indicator.**
- 4. Determine when you are going to administer the instrument or make the observation.**
- 5. Design the instrument.**
- 6. Pilot test and revise the instrument.**



## **Table of Contents**

---

<b>Evaluation – What Is It?</b> .....	7
Program Improvement Evaluation	
Outcome Evaluation	
<b>What Program Are Worth Evaluating?</b> .....	8
Checklist for Deciding Which Programs to Evaluate	
<b>Quantitative or Qualitative?</b> .....	9
<b>Steps in Designing an Outcome-Based Quantitative Instrument to Measure Program Impact</b> .....	9
<b>1. Clearly identify the outcome(s) you want to achieve</b> .....	10-11
Outcome Levels: Learning, Action, Impact	
Examples of Clearly Defined Outcomes	
Knowledge	
Attitude	
Skills	
Behavior	
Practice	
<b>2. Determine how you will measure the achievement of that outcome(s) - indicators and data source</b> .....	12-13
Examples of Indicators and Data Sources	
Knowledge	
Attitude	
Skills	
Behavior	
Practice	
<b>3. Determine the specific details you want to measure under each outcome indicator</b> .....	13-16
Knowledge, Attitude, Skills, Behavior and Practice	
Examples of Specific Details under Outcome Indicators	
Knowledge	
Attitude	
Skills	
Behavior	
Practice	
<b>4. Determine when you are going to administer the instrument or make the observation</b> .....	17-18
Post-Test Only	
Pre and Post-Test	
Post/Post-Test	
Post-Retro Pre	
Post-Test with a Comparison Group	

**5. Design the instrument** ..... 19-30  
    Tips for Designing Questions  
    Knowledge Outcomes  
        True-False  
        Matching  
        Multiple Choice  
        Completion  
    Attitude Outcomes  
    Skill and Behavior Outcomes  
    Observation Checklists for Skills and Behavior Outcomes  
    Intended Behavior and Practice Outcomes  
    Demographic Questions

**6. Pilot test and revise the instrument** ..... 31-33  
    How to Conduct a Pilot Test  
    Item Analysis of Your Question Pool  
        Knowledge Item Analysis  
        Attitudinal Item Analysis  
        Skills and Behavior Item Analysis

**References and Additional Resources on Evaluation**..... 35

**Appendices**

- Outcome Practice Worksheet
- Logic Model Worksheet # 1
- Checklist for the Logic Model and Outcomes
- Logic Model Worksheet #2
- Checklist for Outcomes Indicators and Data Sources
- Space Station Example
  - Logic Model Worksheet #1
  - Logic Model Worksheet #2
  - Output and Outcomes Results

## **Evaluation - What Is It?**

---

Evaluation means different things to different people and takes place in different contexts. Although many definitions have been developed, the one most appropriate for use with community-based organizations is:

***"The systematic collection and analysis of information to determine the worth of a curriculum or program."*** (Alkin 1990)

The word **systematic** means the evaluation must be planned as an integral part of the design or planning process. It is not just an "event" that occurs at the end.

**Worth** can be a measure of:

1. the satisfaction level of participants, leading to **program improvement**, or
2. the **impact of the program** or whether the targeted outcomes have been achieved, otherwise called "**outcome evaluation**."

### **Program Improvement Evaluation**

The data collected from this type of evaluation is primarily used to help improve and implement the program again. It does NOT provide any information about the impact of the program or whether targeted outcomes were achieved.

A program improvement evaluation might include questions such as:

- What did you like best about the program?
- What did you like least about the program?
- What was the most valuable part of the program?
- How would you improve the program?

Or it might include a list of the various sessions or components of the program and ask participants to rate each on a 4- or 5-point scale as the example below.

Please rate the following sessions. (Circle the appropriate number.)

	POOR	FAIR	GOOD	EXCELLENT
Opening Session	1	2	3	4
Gardening with Kids	1	2	3	4
Planning Your Budget	1	2	3	4
Applying Pesticides	1	2	3	4

---

### **Outcome Evaluation**

The data collected from this type of evaluation can be used with community stakeholders and funders to show the impact of a program. Outcome evaluation answers one or more of the following questions:

- Has the program achieved its targeted outcomes?
- How have the participants changed as a result of this program?
- What is the social, economic, civic, or environmental impact of the program?

Both kinds of "evaluation" continue to be important for Extension professionals. However, because of the increased accountability demands, *this field guide will focus primarily on measuring the impact of programs, otherwise called "outcome evaluation."*

## **What Programs Are Worth Evaluating?**

Not all Extension programs need to be formally evaluated. Since developing a good evaluation instrument and analyzing data can be extremely time consuming, be selective in which programs you want to evaluate. Try to select programs where you are spending a lot of contact hours with clientele. These programs have the greatest potential for impacting participants.

Also think about what kind of information your stakeholders are requesting. Do they want to know whether your programs increased knowledge, changed attitudes, or are they more interested in whether the participants developed new skills, changed their behavior, or whether the program resulted in a social, economic, civic, or environmental impact?

Here's a simple checklist to help you decide whether a program is worth evaluating.



### **Checklist for Deciding Which Programs to Evaluate**

- ***Are participants involved in at least two hours of educational programming?***  
It is difficult to show any impact with programs that are shorter than two hours. So you probably want to spend your evaluation efforts on those programs with more contact hours.
- ***Is the program going to be offered again?***  
A one-time program is probably not worth evaluating.
- ***Does the program have clearly stated outcome(s)?***  
You must know the exact outcomes the program is intended to achieve. If you have clearly stated targeted outcomes, it is easier during the program planning process to incorporate activities and components to achieve those outcomes. Clearly stated outcomes are also essential in design of an evaluation instrument.
- ***Are the outcomes measurable?***  
To assess whether a targeted outcome has been achieved, you need outcome indicators that can be observed or measured in some way.
- ***Do your stakeholders care about whether this outcome is achieved?***  
If your stakeholders have a strong commitment to the program, you probably should evaluate whether you are achieving the targeted outcomes. Sometimes your funder will require an evaluation.
- ***Is the program addressing an issue that the community feels is important?***  
It is often beneficial to evaluate programs that address critical issues in your community. If your program is successfully addressing these issues, evaluation data may help you gain additional support.

If you answer “no” to several of these questions, then the program is not probably worth evaluating.



## **Quantitative or Qualitative?**

---

Basically there are two major ways to collect evaluation data: quantitative and qualitative.

**quantitative methodology** - refers to an approach involving the use of numeric-based information that can be measured, compared, and analyzed statistically. This methodology is primarily used as a way to quantify achievement of outcomes as a result of participation in a program. Data are collected through a standardized instrument (i.e., test, survey, and behavior or skill observation checklist). Data are statistically analyzed and results are presented in a numerical format. The advantage of quantitative methodology is that it measures the reactions of a great many people to a limited set of questions, thus facilitating comparison and aggregation of data. Therefore, findings can be generalized.

**qualitative methodology** - refers to an approach that examines, describes, or interprets a program. Qualitative methods make it possible to study a program or issue in more depth and detail than quantitative methods. This methodology is designed to find out what people do, plan to do, think, and feel. There are three primary methods to collect qualitative data: 1) individual or group interviews; 2) direct observations; and 3) examination of written documents such as open-ended items on questionnaires, personal journals/diaries, autobiographies, and program records. Results from a qualitative research study are usually presented in a narrative or pictorial format. Qualitative methods provide vast amounts of detailed information about small numbers of people and cases, while quantitative methods can deal with large numbers of people and limited sets of questions.

<b>Quantitative</b>	<b>Qualitative</b>
Tests*	Interviews
Surveys*	Focus Groups
Questionnaires*	Case Studies
Behavior Checklist*	Naturalistic Inquiry Observation
Skill Checklist*	Testimonials
Product Assessment	Videos
Log	Portfolios
Existing Data/Records	

\* Included in this field guide

## **Steps in Designing a Quantitative Instrument**

---

The rest of this guide will focus on six easy-to-follow steps to design quantitative evaluation instruments to measure whether your educational program has achieved its targeted outcomes. The guide ties into the Logic Model process utilized by many community-based organizations. A copy of the Logic Model is on the inside of the back cover. Logic Model worksheets and checklists are in the Appendix.

**Steps in Designing an Outcome-Based Quantitative Instrument to Measure Program Impact**

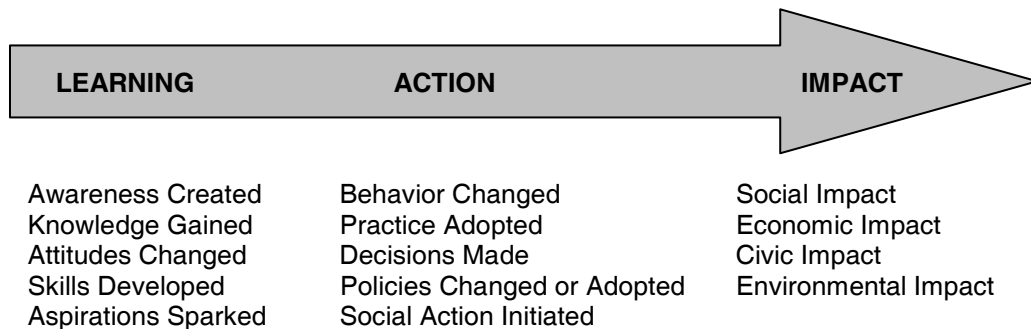
1. Clearly identify the outcome(s) you want to achieve.
2. Determine how you will measure the achievement of that outcome(s) - indicators and data source.
3. Determine the specific details you want to measure under each outcome indicator.
4. Determine when you are going to administer the instrument or make the observation.
5. Design the instrument.
6. Pilot test and revise the instrument.

**1**

**CLEARLY IDENTIFY THE OUTCOME(S) YOU WANT TO ACHIEVE**

An **outcome** is defined as the "change" you would expect to see as a result of participation in a program. In other words, what are the participants going to know, feel, or do differently after participation in the program.

Outcomes can be classified under three major levels: learning, action, and impact.



It is important to be realistic about what outcomes can possibly be achieved by a specific program. Short duration or low concentration programs can usually only achieve **learning outcomes** such as creating awareness, increasing knowledge, changing attitude, or developing skills. These outcomes can be measured immediately following the program.

The greater the number of hours that participants are in the program, the greater the likelihood that higher level "action or impact" outcomes can be achieved. **Action outcomes** include behavior changes, adoption of new practices, decisions made, adoption of a policies, or the initiation of a social action. **Impact outcomes** actually emerge from an action outcome. In other words, if you know what action(s) took place, you can assign a "value" to that action – this is the social, economic, civic, or environmental impact.

**REMEMBER:** Identifying your outcomes should be part of the program planning process. If you can not clearly identify your targeted outcomes, you will have a hard time designing a program to achieve them. Not having clearly defined targeted outcomes is a common evaluation mistake. *A clearly defined outcome(s) is essential in both program planning and program evaluation.*

During the program planning process, as you were completing **Logic Model Worksheet 1**, you should have clearly identified the targeted outcome(s) you expected participants in your program to achieve. Use the **Checklist for the Logic Model and Outcomes** to ensure that you have a clearly defined outcome. A copy of the Logic Model Worksheet #1 and checklist are in Appendix A.

**EXAMPLES OF CLEARLY DEFINED OUTCOMES**

**LEARNING LEVEL**

Knowledge Gained

- Adults will increase their knowledge about using the Food Guide Pyramid as a tool for making healthy food choices.
- Junior Master Gardeners will increase their knowledge about plant growth and development.

Attitude Changed

- Youth will develop a more positive attitude about science, math, and school in general.
- Adults will develop a more positive attitude toward foods that contain products derived from biotechnology.

Skills Developed

- Youth will improve their horseback riding skills.
- Youth will develop their skills (communicating, interacting socially, planning and organizing) through action demonstrations.

**ACTION LEVEL**

Behavior Changed

- Youth will improve their communication skills.

Practice Adopted

- Adults will adopt practices that will decrease the amount of drift when applying pesticides.

**See “Outcome Practice Worksheet” in the Appendix.**

Can you identify the clearly defined outcomes, their level, and what will be measured?

## 2

**DETERMINE HOW YOU WILL MEASURE THE ACHIEVEMENT OF THAT OUTCOME(S) - INDICATORS AND DATA SOURCES**

Now that you have clearly identified the outcomes of your program, it is time to determine how you are going to know whether participants have achieved the outcome(s). In order to measure that achievement you need to identify both the outcome indicators and the data sources. The **Logic Model Worksheet 2** and the **Checklist for Outcome Indicators** are good tools to use for this process. A copy of the Logic Model Worksheet # 2 and checklist are in Appendix B.

**OUTCOME INDICATOR** - the observable, measurable characteristics or change that represents that the participants have achieved the targeted outcomes for the program. An outcome indicator has to be something you can observe or measure. If a condition is not observable and measurable, even though it may relate to the outcome, it is not useful as an indicator. Indicators must be unambiguous. Terms such as "substantial," "acceptable," and "adequate" should not be used.

(United Way, 1996)

**DATA SOURCE** - the instrument or tool that is used to collect data about the outcome indicators. In the case of this field guide, it is an instrument (test, survey/questionnaire, behavior or skill observation checklist) used to collect quantitative data.

- **Tests** – are generally used for knowledge-based questions. Knowledge-based questions offer choices such as correct vs. incorrect or accurate vs. inaccurate. They may ask respondents what they believe is true or factual, or about awareness.
- **Surveys/Questionnaires** – are used to obtain information about what people do, what they have, what they think, know, feel, or want. Survey/questionnaires can be used to measure knowledge, attitudes, skills, behavior, and practice.
- **Observation Checklists** – observation allows the evaluator to watch how an individual behaves or performs a skill in a social setting. Observation checklists can be used to measure skills and behavior.

*Special Note to Researchers:* Remember that sometimes when a person knows they are being observed they act differently. So try to be as unobtrusive in your observations as possible. You may also want to triangulate observation data with other data to increase the rigor of your study.

**EXAMPLES OF INDICATORS AND DATA SOURCES**

**Knowledge Outcome:** Junior Master Gardeners will increase their knowledge about plant growth and development.

**Indicators:** Increase in the number of correct answers on pre/post test.

**Data Source:** Junior Master Gardener Test

**Attitude Outcome:** Youth will develop a more positive attitude about science, math, and school in general.

**Indicators:** Improvement in attitudinal score (from negative to more positive) on pre/post assessment tool.

**Data Source:** Space Station Attitudinal Questionnaire

**Attitude Outcome:** Adults will develop a more positive attitude towards foods that contain products derived from biotechnology.

**Indicators:** Improvement in attitudinal score (from non-accepting to more accepting) on pre/post assessment tool.

A reported increase in the consumption of foods containing products made from biotechnology.

**Data Source:** Biotechnology Questionnaire

**Skill Outcome:** Youth will improve their horseback riding skills.

**Indicators:** Improved performance level when making the horse walk, trot, and canter.

**Data Source:** Observation Checklist

**Skill Outcome:** Youth will develop their skills (communicating, interacting socially, planning and organizing) through Action Demonstrations.

**Indicators:** Ability to perform specific skill indicators.

**Data Source:** Observation Checklist

**Behavior Outcome:** Youth will improve their communication skills.

**Indicators:** Increased frequency of skill set behaviors on post/post assessment.

**Data Source:** Four-Fold Skill Questionnaire

**Practice Outcome:** Adults will adopt practices that will decrease the amount of drift when applying pesticides.

**Indicators:** A reported use of practices (i.e., drift-reducing agents, drift-reducing nozzles) that decrease drift.

**Data Source:** Pesticide Applicators Questionnaire

### 3

## **DETERMINE THE SPECIFIC DETAILS YOU WANT TO MEASURE UNDER EACH OUTCOME INDICATOR**

Before you can start designing your quantitative instrument, you need to determine the specific details of what you want to measure under each outcome indicator. This information will be used to write the items/questions on your instrument.

For example, if your outcome indicator was to increase knowledge about plant growth and development, you need to determine what plant growth and development knowledge you want participants to learn.

You will find that the curriculum, lesson plan, and/or participant handouts you plan to use are very useful in helping to clarify what content will be covered in the program. As you look at these items, ask yourself the following questions based on your targeted outcome.

**Knowledge** – What specific context information (facts, figures, formulas, procedures, categories, etc.) do you want the participants in your program to know? Which content is the most important that participants remember?

**Attitude** – What statements would exemplify a positive or negative attitude toward something? Are there specific actions that would be done by a person with a particular attitude? (i.e., If a person says they would eat foods that contain products derived from biotechnology, you would assume that the person has a positive attitude toward biotechnology.)

**Skills** – What are the specific observed actions that would illustrate that a participant has developed the skill? In many cases, skills are procedural. You have to do certain steps or components in order to master the skill. Besides using a panel of experts in the field or research findings, a good field method for composing the list of specific actions is to conduct a "skill or task analysis." This is accomplished by observing several people doing the particular skill at the mastery level. Make a list of all the specific actions they did. If everyone did a certain action, then that action should be included in your observation checklist.

**Behavior** – What are specific observed actions that would illustrate that the participants had improved their behavior? For example, an after-school program designed to help youth become more successful in school would want to look at the classroom work habits of successful students. Your observation checklist might include behaviors such as: listens well, stays with task until completed, participates in class discussions, follows directions, concentrates on task, completes work promptly, works consistently to ability.

*Special Note:* Behavioral data is more valid when it is actually observed, but that is not always possible. Sometimes program evaluators have to rely on participant self-report of behaviors obtained from questionnaires. For example, if the behavior outcome is that participants eat healthier, actually observing the participant eating healthy foods over a period of time would be impractical. Therefore we have to rely on a self-report through a Food Recall Questionnaire.

**Practice** – What are the specific actions that indicate the participant has adopted a certain practice? Is it practical to observe these actions, or will they have to be done by a participant's self-report?

### **EXAMPLES OF SPECIFIC DETAILS UNDER OUTCOME INDICATORS**

*NOTE:* The specific details listed under each outcome are just examples to illustrate the kind of details that are needed. A comprehensive listing of all the details under the outcome is needed to develop a good quantitative instrument.

**Knowledge Outcome:** Junior Master Gardeners will increase their knowledge about plant growth and development.

**Indicators:** Increase in the number of correct answers on pre/post test.

**Example of Specific Details:**

Sample facts from the chapter on plant growth and development.

- Plants are the beginning of the food chain, providing nourishment for both animals and people.
- The six basic plant parts are: seed, root, stem, leaf, flower, and fruit.
- You can eat all parts of the plant.

**NOTE:** Would also need specific knowledge of the chapters in the Junior Master Gardener manual.

**Attitude Outcome:** Youth will develop a more positive attitude about science, math, and school in general.

**Indicators:** Improvement in attitudinal score (from negative to more positive) on pre/post assessment tool.

**Example of Specific Details:**

Positive statements

- It is fun to do science experiments.
- Learning about space is neat.
- School is fun.

Negative statements

- Science is boring.
- I don't like to do math problems.
- I wish I didn't have to go to school.

**NOTE:** Would need a longer list of statements for the item pool.

**Attitude Outcome:** Adults will develop a more positive attitude toward foods that contain products derived from biotechnology.

**Indicators:** Improvement in attitudinal score (from non-accepting to more accepting) on pre/post assessment tool.

A reported increase in the consumption of foods containing products from biotechnology.

**Example of Specific Details:**

Positive statements

- I would serve genetically modified foods to my family.
- I would eat foods containing products derived from biotechnology research.

Negative statements

- Eating genetically modified foods is harmful to a person's health.
- Genetically modified foods are inferior to foods from plants developed using conventional methods.

**NOTE:** Would need a longer list of statements for the item pool.

**Skill Outcome:** Youth will improve their horseback riding skills.

**Indicators:** Improved performance level when making the horse walk, trot, and canter.

**Example of Specific Details:**

Walk actions

- Rider's ears, shoulder, elbow, hip, and heel in line.
- Rider has appropriate contact with horse.
- Rider sits quietly in saddle.
- Rider in rhythm with the horse.
- Rider's lower legs are steady.

**NOTE:** Would also need specific details for trot, canter, and overall riding skills.

**Skill Outcome:** Youth will develop their skills (communicating, interacting socially, planning, and organizing) through action demonstrations.

**Indicators:** Ability to perform specific skill indicators.

**Example of Specific Details:**

Planning and organizing details

- Topic - suitable length.
- Organizing of content - short, show and tell segments, knew subject in depth, and could answer questions.
- Organizing of content - Are the segments presented in logical order?
- Organizing of content - Are segments explained so that the audience understands "why"?

**NOTE:** Would also need specific details for communicating and interacting socially.

**Behavior Outcome:** Youth will improve their communication skills.

**Indicators:** Increased frequency on skill set behaviors on post/post assessment.

**Example of Specific Details:**

Communication "skill set" behaviors

- Aware of own style of communication (verbal, nonverbal, and listening).
- Understands and values different styles of communication.
- Practices empathy.
- Adjusts own styles of communication to match other's style.
- Communicates of essential information.
- Manages interaction.

**Practice Outcome:** Adults will adopt practices that will decrease the amount of drift when applying pesticides.

**Indicators:** A reported use of practices (i.e., drift-reducing agents, drift-reducing nozzles) that decrease drift.

**Example of Specific Details:**

Management practices to reduce drift

- Chemical - use of a drift-control agent.
- Equipment - select nozzles that produce larger drops, lower the spray pressure, lower the spray boom, increase spray volume.
- Environmental - applying during still or low wind days.



**4****DETERMINE WHEN YOU ARE GOING TO ADMINISTER THE INSTRUMENT OR MAKE THE OBSERVATION**

Before you begin to actually design your quantitative instrument, you need to determine when and how you are going to administer the test, survey, questionnaire, or make the observation. This is called the evaluation methodology. Evaluation methodologies can vary from little rigor to high rigor. An experimental design with an assessment both before and after the program, random assignment of participants, and a control group for comparison is the most rigorous design. This design is primarily used in research or comprehensive evaluation programs.

Because random assignment of participants and control groups is impractical, most Extension educational programs use some type of "quasi-experiment" methodology. The five most common are described below. They vary in the level of rigor as well as the time and effort required. Most Extension field staff would use one of these methodologies.

**POST-TEST ONLY** - The test, survey, questionnaire, or observation is conducted only after the program is completed. Since you are only collecting data once, you can NOT calculate the change that occurred in participants. You can only state that participants have a certain level of knowledge, can perform a skill, or act a certain way. This is not a very rigorous method, but remember schools collect most classroom data this way.

**Examples of impact statements from this method:**

- After the program, 85% of the 45 participants scored 70% or better on the knowledge test.
- Ninety-two percent of the 35 participants were able to design and conduct their own science experiment at the completion of the program.

*NOTE:* You cannot say that participants increase their knowledge, because you have no way of knowing how they would have answered before the program. They may have known all that information already.

**PRE AND POST-TEST** - The test, survey, questionnaire, or observation is conducted both before and after the program. The two scores are then compared indicating whether there is a change in the participants' knowledge, attitudes, skills, and/or behavior. With this method, you are able to say that the change is likely due to the participation in the program. This is more rigorous than post-test only. However, remember that the change in scores could be influenced by other factors (other programs the participants attended, the media, etc.). You could strengthen the rigor of this methodology by testing a similar comparison group (a group that takes the pre/post during the same period, but does not participate in the program).

**Examples of impact statements from this method:**

- Seventy percent of the 100 participants increased their knowledge by an average of eight points (40%).
- Ninety-eight percent of the participants improved their riding skills as measured by a pre/post observation checklist.

*NOTE:* You cannot say participants significantly increased their knowledge or skills unless a statistical analysis of the data, such as a t-test, is conducted.

**POST/POST-TEST** - This is similar to the pre/post-test methodology except that the survey, questionnaire, or observation is conducted immediately following the completion of the program and then again after a period of time (i.e., 30 days, six months). The two scores are then compared, indicating whether the program changed participants' skills and/or behavior.

This design is especially good for self-assessment instruments because it helps eliminate the problem of having participants overestimate their skills and ability, which sometimes occurs with a pre-test. Then once they participate in the program they realize that they didn't know or have the skills they thought they did.

**Example of impact statement from this method:**

- Sixty-five percent of the 35 student improved their classroom work habits as indicated by an increase in behavior frequency on eight of the 14 criteria points.

*NOTE:* Like the pre/post-test, you can not say participants significantly increased their behavior unless a statistical analysis of the data, such as a t-test, is conducted.

**POST-RETRO PRE\*** - The survey or questionnaire is conducted with participants at the end of a program. The instrument is designed to have the participants rate their knowledge, skills, or attitudes at the present time and how they would have rated themselves before the program. Like the pre/post test, the two scores can be compared to indicate whether the program changed participants' knowledge, attitudes or skills. This method is easy to use, but often participants tell you what they think you want to hear - they have been changed by this program - therefore does not have high rigor.

**Example of Post-Retro Pre Questionnaire**

Please rate your ability both before and after you were a teen instructor.										
Key: 1= POOR   2 = NEEDS SOME WORK   3 = OK   4= GOOD   5=EXCELLENT										
	BEFORE TEACHING			AFTER TEACHING						
I would rank my ability to:										
• speak in front of a group	1	2	3	4	5	1	2	3	4	5
• give directions	1	2	3	4	5	1	2	3	4	5
• lead an activity or exercise	1	2	3	4	5	1	2	3	4	5
• keep audience on task	1	2	3	4	5	1	2	3	4	5

**Example of impact statement from this method:**

- Ninety-five percent of the 20 teen instructors improved their planning and organizing skills in these areas: speaking in front of group, giving directions, leading and activity, and keeping an audience on task.

*\*SPECIAL NOTE:* The Post-Retro Pre method should not be used with anyone under age 14. Developmentally, youth have a hard time distinguishing between their knowledge, skills, and attitudes before and after.

**POST-TEST WITH A COMPARISON GROUP** - The test, survey, questionnaire, or observation is conducted with participants after they have completed a program. This data is compared to data that were gathered using the same instrument with a comparison group that did not complete the program. This method is primarily used when both groups (the program participant group and the comparison group) are starting off on an equal footing for the trait or attribute being investigated. When using this method, you really need a statistical analysis, such as a t-test, to show if there was a significant difference between the two groups.

***Example of impact statement from this method:***

- Results of a t-test showed a significant difference in skill performance between the program participants and comparison group. Program participants scored significantly higher on all observation criteria.

## 5

### DESIGN THE INSTRUMENT

Now you are ready to actually write your questions and design the instrument. The design of the instrument takes into account four major issues, which you have worked on in the previous four steps.

- ***What outcome(s) do you want to measure?***  
Knowledge, Attitudes, Skills, Behavior, or Practice
- ***What specific do you want to measure under each outcome?***
- ***What data source do you want to use?***  
Test, Survey/Questionnaire, or Observation Checklist
- ***What evaluation methodology do you want to use?***  
Post-Test, Pre/Post-Test, Post/Post-Test, Post/Retro Pre, or Post-Test with Comparison Group

The procedure for designing an instrument may vary, depending on your answers to the questions above, but all instruments have three common elements:

- Questions
- Response choices
- Demographics that describe the characteristics of your participants

The section on the next few pages provides some helpful tips as you design your instrument.

**NOTE ABOUT QUESTION POOL:** You need to develop more questions than you want in your final instrument, because some questions will be eliminated during the item analysis process of the pilot test step.

**GENERAL TIPS FOR DESIGNING QUESTIONS**

- Ask demographics questions first - this gets the audience engaged in the instrument.
- The first questions should be easy, avoiding controversial topics.
- Make sure questions are related to achievement of the targeted outcome(s).
- Group like questions together – knowledge, attitude, skills, behavior, or practice.
- Keep questions on one subject grouped together.
- Make your questions simple, but do not talk down to your audience.
- Make sure questions have only one thought. To make sure you are only asking one question, do not include the word “**and**” in your questions. (i.e., "How would you rate your financial management knowledge and skills?" - The participants may want to rate their knowledge and skills differently.)
- Avoid questions with the word “**not**” in them.
- Don't quote a question directly from the written curriculum.
- Avoid trick questions.
- Make sure the questions are reasonable and do not invade the respondent's privacy.
- Avoid asking questions that are too precise – such as “how many times did you eat out last month” - use a range instead.
- Avoid using technical jargon or acronyms.
- Remember the ethnic backgrounds of your respondents. Some words have different meanings to different groups.
- Remember the literacy level of your group – you can check the reading level of your instrument in MS Word. (When you run Spelling/Grammar, the last item under readability that shows up is the Flesch-Kincaid Grade Level analysis.)

**KNOWLEDGE OUTCOMES**

If you are using a test to measure knowledge gained, you have basically four different kinds of questions you can use:

1. **True-False** - easy to write and score; tends to promote memorization and encourage a high degree of guessing, so need a large number of questions to gain a true picture of what the participants know.

**TIPS:**

- Write statements that are definitely true or false.
- Make both true and false statements approximately the same length.
- Use approximately equal numbers of true and false statements, but do not place them in a systematic order (i.e., TFTFTF TTFFTT)
- Avoid the following:
  - double-negative statements
  - complex sentences
  - unintended verbal clues
  - indefinite degrees (i.e., large, long time, regularly)
  - absolutes (never, only, always)

2. **Matching** - easy to write and score; ideal for measuring associations between facts; tends to emphasize memorization.

**TIPS:**

- Put descriptions, longer phrases, or statements in the first column and the shorter options, words or symbols in the second column.
- There should be at least three more options than descriptions – this makes it harder for respondents to eliminate options.
- In directions, specify whether the basis for matching and whether options can be used more than once.

3. **Multiple Choice** - considerable versatility in measuring outcomes; designed to make participants discriminate among options that vary in degree of correctness, thus a higher level of cognitive thinking than T/F questions; time-consuming to write and have to be carefully written.

**TIPS:**

- The question should clearly state the problem and keep response options short.
- Be sure wrong choices are plausible.
- Rotate the position of the correct choice from question to question.
- Use the option “none of the above” sparingly.
- Avoid using the option “all of the above.”

**NOTE:** Remember, multiple-choice questions can also be used to measure higher cognitive levels such as comprehension, application, analysis, synthesis, or evaluation. To measure higher cognitive levels you might want to use one of the following approaches:

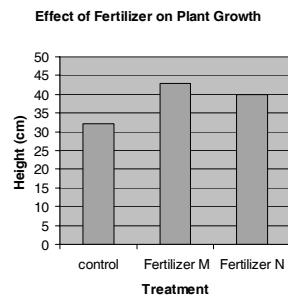
- Use pictures, drawing, graphs, tables, etc. that require the respondent to think at the application or above level.
- Use analogies that demonstrate relationship among terms.
- Require the application of previously learned principles or procedures to new situations.

**Example of a Higher Cognitive Level Question**

1. In a science experiment, one bean seed was placed in each of nine pots with the same amount of soil. The first three pots were watered with water only – no fertilizer. The next three pots were watered with fertilizer M. The last three pots were watered with fertilizer N. At the end of five weeks the plants were measured.

Looking at the graph, which treatment produced the tallest plants?

- a. no fertilizer
- b. fertilizer M
- c. fertilizer N
- d. the control



4. **Completion** - relatively easy to write, but sometimes hard to score because there can be more than one defensible answer; restriction of an answer to a few words tends to measure the recall of specific facts, names, places, and events as opposed to more complex items.

**TIPS:**

- Statements should require a single-word answer.
- Avoid statements that may be logically answered by several terms.
- Omit only key words: don't eliminate so many elements that the statement doesn't make sense.
- Be sure the answer that is required is factually correct.

**NOTE:** To show that participants in your program *increase their knowledge*, you need to administer both a pre-test and a post-test and compare the scores.

**ATTITUDE OUTCOMES**

If you are trying to measure change in attitudes, you are basically looking at three types of attitudinal statements:

1. **Cognitive Items** - express beliefs about the attitudinal object.  
Belief: Genetically modified foods are inferior to foods from plants developed using conventional methods.
2. **Affective Items** - express a feeling toward the attitudinal object.  
Feeling: I like to do science experiments.
3. **Conative Items** - express behavioral intention or behavioral preference with regard to the object. *Would* items express a personal behavioral intention to do something. *Should* items express a behavioral preference for social action.  
Would: I would serve genetically modified foods to my family.  
Should: The government should ban the production of genetically modified foods.

Asking a direct question where the respondent can only answer yes or no (i.e., Do you think genetically modified foods are bad?) is not recommended. It is difficult for an educational program to change a person's attitude from an absolute "no" to an absolute "yes." Instead, have respondent rate numerous statements about the topic on a scale such as: strongly disagree, disagree, undecided, agree, and strongly agree. See the example on the next page.

The mean (average) rating of those statements indicates the respondent's attitude toward the topic. Then by using a pre/post test you can compare the pre mean to the post mean to determine if there was any movement in attitude.

**Examples of Attitude Questions and Response Choices**

Please circle the letter that corresponded to your feelings about each of the statements below.

	STRONGLY DISAGREE	DISAGREE	UNDECIDED	AGREE	STRONGLY AGREE
1. I would serve genetically modified foods to my family.	a	b	c	d	e
2. Eating genetically modified foods is harmful to a person's health.	a	b	c	d	e
3. Genetically modified foods are inferior to foods from plants developed using conventional methods.	a	b	c	d	e
4. I would eat foods containing products derived from using biotechnology research.	a	b	c	d	e

NOTE: Scoring for this test would be as follows: a = 1, b =2, c =3, d = 4, and e = 5. The lower the mean (average) of the respondent's score, the more negative their attitude toward genetically modified foods.

But remember, you need to use reverse coding for negatively worded statements. Scoring for negatively worded questions would be: a = 5, b =4, c =3, d = 2, and e = 1. So, if a respondent answered "e" on question 2, you would score it as a "1" not a "5".

**Other Attitudinal Response Choices** - Response choices vary depending on the outcome being measured and how the question is worded.

When using Likert-like scales, it is important to make sure that the scales are balanced. Your scale should have an equal number of both positive and negative options. The scale used depends on:

- how the question is asked
- the amount of differentiation that is possible and desirable, and
- the respondents' age and capacity to answer.

Below are some examples of scales that can be used with attitudinal questions.

STRONG DISAGREE  
DISAGREE  
NOT SURE  
AGREE  
STRONGLY AGREE

STRONGLY OPPOSE  
OPPOSE  
NEUTRAL  
FAVORABLE  
STRONGLY FAVORABLE

EXTREMELY UNFAVORABLE  
UNFAVORABLE  
NEUTRAL  
FAVORABLE  
EXTREMELY FAVORABLE

DISAGREE A LOT  
DISAGREE A LITTLE  
AGREE A LITTLE  
AGREE A LOT

DEFINITELY FALSE  
FALSE  
NOT SURE  
TRUE  
DEFINITELY TRUE

NOT IMPORTANT  
RELATIVELY UNIMPORTANT  
MODERATELY IMPORTANT  
HIGHLY IMPORTANT  
DON'T KNOW

Sometimes you may want to use a scale with more discrimination points.

*Example:* For each statement below, circle the number that indicates your degree of favorableness towards the statement.

EXTREMELY UNFAVORABLE			NEUTRAL			EXTREMELY FAVORABLE		
1	2	3	4	5	6	7	8	9

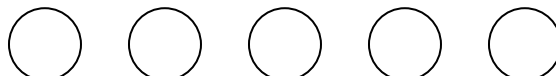
*Example:* For each question below, circle the “X” on the continuum indicating how strongly you favor or oppose the statement.

What is your attitude towards .....

STRONGLY OPPOSE			NEUTRAL			STRONGLY FAVOR		
X	X	X	X	X	X	X	X	X

For young children with limited reading ability, you may want to use smiling face graphics.

*Example:* I like to do science experiments.



**SKILL AND BEHAVIOR OUTCOMES**

Skills and behaviors are usually measured in one of two ways:

1. **Asking participants** whether they learned a skill or are practicing a certain behavior. This self-assessment data is usually collected through a pre/post or a post/retro pre questionnaire.
  - Skill questions tend to ask respondents to rate their ability or the confidence about their ability.
  - Behavior questions usually ask respondents about the frequency of their behavior (how often they do something).
2. **Observing the participants** actually performing the skills or doing the behavior. Observation data is usually collected using a pre/post or post/post observation checklist. Sometimes you can ask parents, teachers, or someone who interacts with the respondent frequently whether they have noticed any changes in behavior.

Observing behavior is more rigorous than self assessment. However, a combination of a self-assessment followed by an observation from a parent, teacher, etc. can increase the rigor. This is called “triangulation,” collecting data from more than one source.



**Examples of Skill Questions**

Please circle the letter that best describes your ability to do the following things:

	CANNOT DO IT	MUCH HELP NEEDED	MAYBE	LITTLE HELP NEEDED	CAN DO IT
1. Finding someone's e-mail address on the web.	a	b	c	d	e
2. Using HTML to design a simple web page.	a	b	c	d	e
3. Adding motion and sound to your web page.	a	b	c	d	e

---

Please circle the phrase that best describes your comfort level in doing the following things:

1. How comfortable do you feel about your ability to lead a small group activity?

VERY UNCOMFORTABLE  
 SOMEWHAT UNCOMFORTABLE  
 SOMEWHAT COMFORTABLE  
 VERY COMFORTABLE

---

Please rate your ability both before and after you were a teen instructor.

Key: 1 = POOR 2 = NEEDS SOME WORK 3 = OK 4 = GOOD 5 = EXCELLENT

	Before teaching					After Teaching				
I would rank my skills as follows:										
• communication skills	1	2	3	4	5	1	2	3	4	5
• leadership skills	1	2	3	4	5	1	2	3	4	5

**Other Skill Response Choices:** Below are some examples of scales that can be used with skill questions.

NOT VERY PREPARED  
 SOMEWHAT PREPARED  
 MODERATELY PREPARED  
 WELL PREPARED

VERY UNCOMFORTABLE  
 SOMEWHAT UNCOMFORTABLE  
 SOMEWHAT COMFORTABLE  
 VERY COMFORTABLE

POOR  
 NEEDS SOME WORK  
 OK  
 GOOD  
 EXCELLENT

CAN NOT DO IT  
 MUCH HELP NEEDED  
 MAYBE  
 LITTLE HELP NEEDED  
 CAN DO IT

**Examples of Behavior Questions**

Directions: Place an "X" in the box that best corresponds to how often you did what is described in the last 30 days.

	NEVER	RARELY	SOMETIMES	OFTEN	ALWAYS
1. When I have a problem, I first figure exactly what the problem is.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. I try to get all the facts before trying to solve a problem.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. When solving a problem, I do the first thing that comes into my head.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please think of the past week. Using the scale below, please circle the best answer that tells how often you used the following discipline strategies.

**0 = NEVER    1-2 = ONCE OR TWICE    3-5 = 3-5 TIMES    5+ = MORE THAN 5 TIMES**

In the past week.....

I caught my child doing something right and told him/her.	<b>0</b>	<b>1-2</b>	<b>3-5</b>	<b>5+</b>
I caught my child doing something wrong and told him/her.	<b>0</b>	<b>1-2</b>	<b>3-5</b>	<b>5+</b>
I spanked my child.	<b>0</b>	<b>1-2</b>	<b>3-5</b>	<b>5+</b>
I used a time out.	<b>0</b>	<b>1-2</b>	<b>3-5</b>	<b>5+</b>

Directions: Please circle the number corresponding to the action you have observed from participants in this program during the last week.

	NEVER	SELDOM	SOMETIMES	USUALLY	ALWAYS
How often do participants in this program:					
1. act polite and courteous (i.e., not talking with someone else is talking)	1	2	3	4	5
2. use "put downs"	1	2	3	4	5
3. use threats, physical force, or anger when settling disputes or disagreements	1	2	3	4	5

**NOTE:** You can also ask the respondent to give you an exact number as the example below.

How many days during the week do you participate in physical activities? \_\_\_\_\_ days (answers can range from 0 to 7 days)

When you participate in physical activities, what is the average length of time in minutes that you spend? \_\_\_\_\_ minutes

**Other Behavior Response Choices:** Below are some examples of scales that can be used with behavior questions.

NOT AT ALL  
SOME (LITTLE)  
A LOT  
A GREAT DEAL

HARDLY EVER  
RARELY  
SOMETIMES  
OFTEN  
VERY OFTEN

NEVER  
ALMOST NEVER  
SOMETIMES  
FAIRLY OFTEN  
VERY OFTEN  
ALWAYS

NEVER  
LESS THAN ONCE A WEEK  
ONCE A WEEK  
2 – 3 TIMES A WEEK  
4 – 5 TIMES A WEEK  
DAILY  
SEVERAL TIMES A DAY

NOT AT ALL  
ONCE OR TWICE  
3 TO 5 TIMES  
6 OR MORE TIMES

NEVER  
OCCASSIONALLY  
SOME OF THE TIME  
MOST OF THE TIME  
ALL OF THE TIME

**OBSERVATION CHECKLISTS FOR SKILLS AND BEHAVIOR OUTCOMES**

As was mentioned earlier, observing participants actually performing the skill or doing the behavior is a more rigorous way to measure improvement in skills or behaviors. Usually this is done with an observation checklist.

Remember under **Step 3** we determined the specific details we wanted to measure. Here’s a brief summary of how you determine what should be included on an observation checklist for skills and behavior.

**Skills** – In many cases, skills are procedural. Certain steps or components need to be completed in order to master the skill. Besides using a panel of experts in the field or research findings, a good field method for composing the list of specific actions is to conduct a "skill or task analysis." This is accomplished by observing several people doing the particular skill at the mastery level. Make a list of all the specific actions they did. If everyone did a certain action, then that action should be included in your observation checklist.

**Behavior** – Developing a checklist to observe certain behaviors is a little more difficult. A panel of experts is probably the best way to list the specific observed actions that would illustrate that the participants had improved a certain behavior. For example, an after-school program designed to help youth become more successful in school would want to look at the **classroom work habits of successful students**. Your observation checklist might include behaviors such as: listens well, stays with task until completed, participates in class discussions, follows directions, concentrates on task, completes work promptly, works consistently to ability.

Observation checklists are usually done using a pre/post methodology. However, you can observe behavior periodically throughout the program if you want to know how the participants are progressing.

*SPECIAL NOTE:* Remember that sometimes when a person knows they are being observed they act differently. So try to be as unobtrusive in your observations as possible. You may also want to triangulate observation data with other data to increase the rigor of your study.

**Examples of Skill Observation Checklists**

*NOTE:* The examples below are only part of an entire checklist.

**Horsemanship Checklist** (Note: This checklist would require someone who has expertise in horsemanship to serve as the observer.)

Key: 1 = POOR 2 = SOMEWHAT 3 = OK 4 = GOOD 5 = EXCELLENT

Skills displayed during trotting.

1. Are the rider's ear, shoulder, elbow, hip, and heel in line?	1	2	3	4	5
2. Does the rider have the appropriate amount of contact with the horse?	1	2	3	4	5
3. Does the rider sit quietly in the saddle?	1	2	3	4	5
4. Is the rider in rhythm with the horse?	1	2	3	4	5
5. Are the rider's lower legs steady?	1	2	3	4	5

**Action Demo Checklist**

Key: 1 = POOR 2 = NEEDS IMPROVEMENT 3 = OK 4 = GOOD 5 = EXCELLENT

Presenting the demonstration

1. Does 4-H'er encourage audience involvement by having a "hands-on" activity?	1	2	3	4	5
2. Does 4-H'er speak directly to the audience?	1	2	3	4	5
3. Does 4-H'er show evidence of practice and experience?	1	2	3	4	5
4. Does 4-H'er tell about what they have learned through this 4-H project?	1	2	3	4	5

**Example of a Behavior Checklist**

Directions: Observe the student and rate the frequency of behavior. First rating should be done near the beginning of the program. The second rating should be done near the end of the program.

Rating Key: 1= NEVER 2 = OCCASSIONALLY 3= 50% OF TIME 4 = FREQUENTLY 5= ALWAYS

Behavior	1 <sup>st</sup> Rating	2 <sup>nd</sup> Rating
Listens well		
Takes turn in a game, etc.		
Participates in activities		
Stays with task until completed		
Participates in class discussions		
Respects rights of others		

**INTENDED BEHAVIOR OR INTENDED ADOPTION OF PRACTICE OUTCOMES**

Sometimes, time and money prevent us from actually measuring behavior or adoption of a practice after a program is over. In those cases, you might want to measure the participants' intentions. Psychology suggests that knowing a participant's intention is a good predictor of behavior. You can use a **post survey** to measure intentions.

When trying to measure intention, don't phrase the question so that they can only answer "yes" or "no." Most participants don't want to admit they don't plan to do something you just worked hard to convince them to do, therefore skewing your data. If you give respondents a range of answers to choose from, they will be more likely to answer honestly. (Kiernan, 2001)

**Examples of Intended Behavior and Adoption of Practice**

Please circle the answer indicating which action step(s) you currently do or whether plan to do them in the future. This survey is completely anonymous, so please answer questions honestly.

1. Talk with my child about ways to say "no" or avoid situations where someone offers them tobacco, alcohol, or other drugs.

HIGHLY UNLIKELY      UNLIKELY      MAYBE      HIGHLY LIKELY      CURRENTLY DO

2. Set and enforce clear rules about not using tobacco, alcohol, and other drugs.

HIGHLY UNLIKELY      UNLIKELY      MAYBE      HIGHLY LIKELY      CURRENTLY DO

---

As a result of this program, how willing are you to switch to high-nutrient feed for your dairy herd?

NOT WILLING      RELUCTANT      SOMEWHAT WILLING      QUITE WILLING      VERY WILLING

**Other Intended Behavior or Practice Response Choices:** Below are some examples of scales that can be used to find out people's intentions.

DEFINITELY NO  
 PROBABLY NO  
 POSSIBLY NO  
 UNCERTAIN  
 POSSIBLY YES  
 PROBABLY YES  
 DEFINITELY YES

DEFINITELY NO  
 PROBABLY NO  
 UNCERTAIN  
 PROBABLY YES  
 DEFINITELY YES

NOT WILLING  
 RELUCTANT  
 SOMEWHAT WILLING  
 QUITE WILLING  
 VERY WILLING

HIGHLY UNLIKELY  
 UNLIKELY  
 MAYBE  
 HIGHLY LIKELY

### DEMOGRAPHICS QUESTIONS

Not everyone in your program will gain the same benefits. Sometimes this difference may be due to age, gender, interest level, or previous experience or training. For example, if you are conducting a leadership development program and some of your participants have had previous training, then you need to know that. It is highly likely that they will not learn as much as those new to the topic. That is why it is important to collect some demographic data about your participants.

What demographic data you collect is very dependent on your audience and the topic area of the program. Demographic questions add to the length of your test, or survey/questionnaire, and it is important to keep it short. Only ask about the attributes you really think are important to know. Also be careful about privacy. Instead of asking what someone's exact gross income is, it is better to give the ranges.

The following is a list of background questions you might include in your instrument:

- Gender
- Age
- Ethnicity
- Education
- Marital status
- Income
- County
- Occupation
- Size of farm/business operation
- Employment status/position
- Certifications held
- Prior knowledge of a topic
- Number of children or family size

Always make sure your demographics stay current with the times (see the following example to use when addressing marital status).

1. What is your marital status? (circle one)
  - a. Single
  - b. Living with significant other or partner
  - c. Currently Married
  - d. Divorced or separated
  - e. Widowed
2. What is the highest level of education that you have completed? (circle one)
  - a. Less than high school
  - b. High school graduate
  - c. Technical school or some college
  - d. College graduate or beyond

(Barkman, Machtmes, 2002)

**6****PILOT TEST AND REVISE THE INSTRUMENT**

A pilot test is really a trial run of your instrument to see how it works. A pilot test helps you increase the accuracy and consistency of your instrument. Before pilot testing your instrument on an audience, it is always good to let another person who is not familiar with your program review the instrument. They can often find jargon and other words that may not be understood by the audience. Make these changes and then you are ready to pilot test.

**HOW TO CONDUCT A PILOT TEST**

The instrument should always be pilot tested on 10 to 15 people who are representative of the target audience for the program. These should not be people who may later be participants in the program.

During the pilot test you want to discover the answers to the following questions:

- Do the pilot participants find the instrument easy to use?
- Do the pilot participants find the response choices adequate and sufficient? Do they understand what each response choice means?
- Do the pilot participants understand the meaning of each question? Are there any words they don't understand? Are there any words they think are misleading or have double meaning?
- Do they have any suggestions that would improve the questions or the format of the instrument?

As you look at how pilot participants answered the questions, you should be able to answer the following questions:

- Did pilot participants find the questions to mean what they are intended to mean?
- Do the answers to the questions provide information about the achievement of program outcomes?
- How consistent was the information obtained from the instrument?
- How accurate was the information obtained from the instrument?

*NOTE:* If your pilot participants had difficulty understanding a lot of the questions, the response choices, or the format, you should redesign the instrument and pilot again. You need to correct the major problems with the instrument before attempting to do an item analysis.

**ITEM ANALYSIS**

After you have administered the instrument to the pilot participants, you should conduct an item analysis to determine which questions you should keep or eliminate. This process is different depending on what you are measuring: knowledge, attitudes, skills, behavior, or practice.

Once you have administered the test, you need to record all the scores on a piece of notebook or graph paper. Below are examples of how to conduct a quick field item analysis on different types of questions without even having to use a computer.

**Knowledge Item Analysis**

- Record the number of correct answers to each question.
- Eliminate questions where 50% or more of the pilot participants got the answer right.

Assume there were 16 pilot participants, so in this example: questions 1, 4, 5, 9, and 10 would be eliminated.

You eliminate these questions because if your participants already know this information, your program is not going to help them gain that knowledge. You want questions that illustrate a gain in knowledge from the pre-test to the post-test.

Q1.	xxxxxxxxxxxx
Q2.	xxx
Q3.	xx
Q4.	xxxxxxxx
Q5.	xxxxxxxxxxxx
Q6.	x
Q7.	
Q8.	xx
Q9.	xxxxxxx
Q10.	xxxxxxxxxxxx
Q11.	
Q12.	
Q13.	xx
Q14.	xxx
Q15..	xxxxx

**Attitudinal Item Analysis**

- Record the number of responses to each question.
- Eliminate questions where most of the responses were grouped in the middle (around b, c, d).

In this example: questions 2, 9, 10, 13, and 15 would be eliminated.

You eliminate these questions because they don't show that participants have a clear negative or positive attitude. You want questions that tend to polarize attitudes, that way you have some chance for movement after the program.

	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>	<b>e</b>
Q1.	8	2	-	1	4
Q2.	1	3	6	5	1
Q3.	6	1	1	1	6
Q4.	10	1	2	-	-
Q5.	8	1	-	-	5
Q6.	7	2	1	1	4
Q7.	5	2	-	2	6
Q8.	6	2	-	2	5
Q9.	-	4	7	4	-
Q10.	-	2	10	3	-
Q11.	8	2	-	1	4
Q12.	8	1	-	1	5
Q13.	2	2	7	2	2
Q14.	9	-	-	-	6
Q15.	1	2	9	2	1

**SPECIAL NOTE FOR RESEARCHERS:**  
 In order to establish the psychometrics for your instrument, you need to run a statistical test for reliability. A Cronbach alpha is one way to measure reliability. To measure reliability using SPSS, request a reliability analysis under the heading of statistics. Reading reliability results requires some experience. Please seek help to understand this analysis. A factor analysis allows a researcher to examine correlations among variables and should be used when researchers are developing instruments. Please seek help when attempting a factor analysis. Validity may be measured in several different formats.



### **Skills and Behavior Item Analysis**

Although the field method for item analysis for survey/questionnaires and observation checklists are not as formal as those for knowledge and attitude questions, the instruments still need to be pilot tested. Use the questions on page 31 as a guide. You may need to change the wording on questions or maybe even change the number of response choices.

With observation checklists, it is imperative to pilot test – observe a real audience.

- Was it easy to observe the skills or behavior listed on the checklist? If not, what needs to be changed?
- Were there skills and behaviors you observed that need to be added to the checklist?
- Could you differentiate between the various response choices? Were choices too far apart or too close in nature?

With intended behavior survey/questionnaires, be sure to check how respondents felt about the wording of the questions? Did they feel reluctant to answer? Were the response choices adequate?



## **References and Additional Resources on Evaluation**

- Alkin, Marvin, C. (1990). Debates on Evaluation. Newbury Park, CA: SAGE Publications.
- Alter, C. & Eagan, M. (1997). Logic modeling: A tool for teaching critical thinking in social work practice. Journal of Social Work Education, 33 (1).
- Alter, C. & Murty, S. (1997). Logic modeling: A tool for teaching practice evaluation. Journal of Social Work Education, 33 (1).
- Barkman, S.J. & Machtmes, K.L. (2002). Designing Quantitative Instruments. West Lafayette, IN: Purdue University.
- Barkman, S.J., & Machtmes, K.L. (2000). Evaluating the impact of youth development programs user guide (4-H 913). West Lafayette, IN: Purdue University.
- Barkman, S.J. & Machtmes, K.L., et. al. (1999). Evaluating 4-H curriculum through the design process; pilot testing and collecting data for the 4-H national jury review process (4-H 898). West Lafayette, IN: Purdue University and The Ohio State University.
- Doran, D., Chan, F., & Tamir, P. (1998). Science educator's guide to assessment. Arlington, VA: National Science Teachers Association.
- Fink, A. (1995). The survey kit (Vols. 1-9). Thousand Oaks, CA: SAGE Publications.
- Gall, M.D., Borg, W.R., & Gall, J.P. (1996). Educational research: An introduction (6<sup>th</sup> ed.) New York: Longman Publishers.
- Kiernan, N.E. (2001) Measuring Intentions: Tipsheet 50. University Park, PA, Penn State Cooperative Extension. <http://www.extension.psu/evaluation/tipsheets/measuringintentions.pdf>.
- Kubiszyn, T. & Borich, G. (1996). Educational testing and measurement: Classroom application and practice (5<sup>th</sup> ed.) Harper Collins College Publishers.
- Morris, L.L., Fitz-Gibbon, C.T., & Lindheim, E. (1987). How to measure performance and use tests. Newbury Park, CA: SAGE Publications.
- Ostlund, K.L. (1992). Science process skills: Assessing hands-on student performance. Menlo Park, CA: Addison-Wesley Publishing Company.
- Taylor-Powell, E. (1999, June). Logic model notes. Paper presented at the University of Wisconsin's providing leadership for program evaluation, Madison, Wisconsin.
- United Way of America. (1996). Measuring program outcomes: A practical approach. Arlington, VA: United Way of America.



## **APPENDICES**

**Outcome Practice Worksheet**

**Logic Model Worksheet # 1**

**Checklist for the Logic Model and Outcomes**

**Logic Model Worksheet #2**

**Checklist for Outcome Indicators and Data Sources**

**Space Station Example**

Logic Model Worksheet #1

Logic Model Worksheet #2

Output and Outcome Results



## Outcome Practice Worksheet

---

Directions: Please place an “X” beside the outcomes that you think are clearly stated. Then write the outcome level (learning, action, impact) and what will be measured in the space below.

**Example:**   X   Youth will improve their horseback riding skills.  
 Level:   Action – skills improved  



- 1. Adults will develop a more positive attitude toward foods that contain products derived from biotechnology.  
 Level: \_\_\_\_\_
- 2. Increase the number of schools that offer school enrichment programs.  
 Level: \_\_\_\_\_
- 3. Adults will adopt practices that will decrease the amount of drift when applying pesticides.  
 Level: \_\_\_\_\_
- 4. Parents will increase their use of positive parenting techniques.  
 Level: \_\_\_\_\_
- 5. Youth will improve their classroom behavior.  
 Level: \_\_\_\_\_
- 6. Program participants will increase their involvement in community affairs.  
 Level: \_\_\_\_\_
- 7. Master gardeners will increase their knowledge about garden management.  
 Level: \_\_\_\_\_
- 8. Pregnant teen mothers will eat healthier and not consume alcohol, tobacco, and other drugs.  
 Level: \_\_\_\_\_
- 9. Adults will improve their credit rating.  
 Level: \_\_\_\_\_
- 10. The educator will create an awareness and knowledge of gang-related symbols.  
 Level: \_\_\_\_\_





# Logic Model Worksheet #1

Name of program \_\_\_\_\_

INPUTS	OUTPUTS		OUTCOMES - IMPACT		
	Activities	Participation	LEARNING	ACTION	IMPACT
<p>What do we need to achieve our goals?</p> 	<p>What do we have to do to ensure our goals are met?</p>	<p>Who needs to</p> <ul style="list-style-type: none"> <li>- participate?</li> <li>- be involved?</li> <li>- be reached?</li> </ul> 	<p>What do we think the participants will know, feel, or be able to do after participation in the program?</p>	<p>How do we think the participants will behave or act differently after participation in the program?</p>	<p>What kind of impact can result if the participants behave or act differently after participation in the program?</p>

Adapted from E. Taylor-Powell, 1999

# Outcome Indicator Worksheet #2

Name of program - \_\_\_\_\_

OUTCOMES	INDICATOR(S) (1 or more per outcome)	DATA SOURCE
Outcome Level		
Outcome Level		
Outcome Level		

# Logic Model Worksheet

Name of Program - Space Station Indiana

INPUTS	OUTPUTS		OUTCOMES - IMPACT		
	Activities	Participation	LEARNING	ACTION	IMPACT
<p><b>Volunteers</b> At least 3 volunteers per day in SI</p> <p><b>Parent Helpers</b> 1 parent hired for every 5 youth in program (1:5)</p> <p><b>Project Director</b> 1 FTE in summer months and .75 FTE in other months</p> <p><b>School Corp.</b> helps recruit students and provide bus transportation for SI participants. Provide classroom space for SI program.</p> <p><b>CES</b> coordinates program and supervises Project Director</p>	<p><b>5 MAJOR COMPONENTS</b></p> <p><b>Summer Intensive Program (SPI)</b> 5 days (11 -4 pm) 4 weeks 100 contact hours</p> <p><b>Follow-Up Group Meetings</b> once a month through rest of year (2 - 4 hours) 38 contact hours</p> <p><b>School Enrichment</b> once a week, 45 minutes during school year 28.5 contact hours</p> <p><b>Parent Helpers</b> hire and train parents as SI classroom helpers</p> <p><b>Parenting Classes</b> once a week for 6 weeks</p>	<p><b>First - 3rd grade students</b> who failed the ISTEP test or recommended by teacher</p> <p><b>Parent Helpers</b> will work 5 days/week for 4 weeks (10 - 4 pm) - need to be parents of participants. Must have 1 adult for every five youth (Shuttle Crews)</p> <p><b>Parents of SI</b> participants will attend parenting classes</p> <p><b>Volunteers</b> Jr. Leaders and other community people</p>	<p>Improve the <b>science and math literacy</b> of participants</p> <p>Develop a <b>more positive attitude</b> toward science, math, and school in general in 80% of the participants</p>	<p>Improve classroom <b>behavior</b> of SI participants</p> <p>Improve the <b>behavior</b> of the parents towards their children</p>	