

C. INSTITUTION REVIEW BOARD (IRB) PROCESS

If you are working with an academic institution, your evaluator may be involved in a review process with the college or university's Institutional Review Board (IRB). It is important to keep in mind that the IRB process generally takes several weeks to complete and may add time to the start-up of the project.

The OCMH expects grantees to perform direct service within the first quarter of project funding. Therefore, you are encouraged to simultaneously apply for an IRB when you apply for OCMH funding. If it is later determined that you will not use the IRB there will be no detriment to the OCMH funded project.

An IRB is a federally-recognized committee authorized to review research projects and ensure that procedures comply with ethical standards and established to protect the rights and welfare of the human subjects involved in research activities such as program evaluation. IRB committees review proposed research plans to ensure that researchers educate potential participants about the purposes, risks, and benefits of the research, inform them of the voluntary nature of their participation, their right to withdraw at any time and their right to receive information about the research results. IRB committees also make sure that written consent has been received from participants as well as ensuring a plan is in place to keep participant information confidential.

Many colleges, government offices, hospitals, and research agencies have established IRBs. IRB approval is generally not required for project/program evaluations. In rare instances it may be required, especially with some types of federal funding. Evaluations may also require IRB approval when the evaluator intends to share the results outside of the program being evaluated, such as in published articles or journals. Some typical evaluations that might require IRB approval are: Evaluations that use sensitive health information, such as information from medical records, and Evaluations where you contract or work with a faculty member or student from any academic setting. More information about IRBs and research with human subjects can be found at: <http://www.hhs.gov/ohrp/>

D. PROTECTING PARTICIPANT RIGHTS IN EVALUATION

In all studies involving human subjects it is crucial to think through how the project ensures that the rights of individual participants are protected. Even if a study is not research per se, the program staff and their evaluator should ensure that their practices respect individual's rights. All participants should be guaranteed the following –

1. Informed consent – Participants should be made aware that their participation in the study is wholly voluntary and they can decline to participate at any time or in any aspect of the project. Their decision to decline in no way impacts their receipt of program services or their relationship with the delivering organization.
2. Confidentiality – Any information provided by participants will be kept confidential and only used for the purposes of the evaluation. No specific data will be attributed to

individuals and no identifying information would be presented in any reports. This also extends to ensuring the participants' privacy while they are providing information for the study. When sign in sheets, logs and other tracking documents are used, full names should not be recorded and/or reported. Participant information should be kept in a locked cabinet.

3. Projects must have authorization to photograph video tape and/or use other electronic formats of participants. Written release forms must be signed and dated. Parents/guardians need to sign for minors. **Do not submit participant photographs, video recordings and/or other digital images to the OCMH unless requested by the OCMH and if requested provide required**

Federal privacy laws (Section adapted from Evaluation Ethics. Wilder Foundation, February, 2009.)

There are a few federal laws that serve to protect the privacy of research participants, including the Federal Educational Rights and Privacy Act (FERPA) and the Protection of Pupil Rights Act (PPRA). Even if no elements of your current evaluation are subject to any federal privacy laws, but it is good to be aware that such laws do exist. The following briefly describes the policies that could potentially affect evaluation efforts of the current project or future projects.

Health Insurance Portability and Accountability Act (HIPAA) – HIPAA is a federal law enacted in 1996 designed to protect the privacy and security of health information. This could include information collected about chemical health or mental health. For information about HIPAA, visit this link: <http://www.hhs.gov/ocr/hipaa/>.

VI. LOGIC MODEL AND PROGRAM OUTCOME MEASUREMENT

The OCMH plans to fund projects that develop logic models that clearly show the link between program inputs, activities, outputs, outcomes, and outcomes indicators. These types of programs help the OCMH to accomplish parts of its mission. The logic model is a planning tool used to clarify and illustrate what your project intends to do and what it hopes to accomplish and impact. More specifically, a logic model:

- summarizes key program elements;
- explains rationale behind program activities;
- shows the links between program inputs, activities, outputs, outcomes, and outcome indicators;
- clarifies intended outcomes; and
- provides a communication tool.

The logic model can be considered a map that you develop to clarify and communicate what your project intends to do and its presumed impact. It consists of program inputs, activities, outputs, outcomes, and outcome indicators. These components are linked together to form the building blocks of a logic model. A logic model template is shown in Figure 1. Figure 1 provides an example of a well-constructed logic model for a prevention of chronic diseases and conditions.

FIGURE 1: GENERIC LOGIC MODEL

Inputs	Activities	Outputs	Outcomes	Outcome Indicator(s)
<p>What is used to deliver services</p> <p>Resources dedicated to or consumed by the program: ✓ Staff (administrative, medical and program) & staff Time ✓ Volunteers & volunteers' time ✓ Facilities ✓ Equipment & Supplies (clinical tests)</p> <p>Constraints on the program: ✓ Laws ✓ Regulations ✓ Funder's requirements</p>	<p>Things done to, for, or with the target population intended to lead to the desired change</p> <p>What the program does with inputs to fulfill its mission: ✓ Conduct a pre-diabetes prevention program, blood pressure reduction, breast cancer screening programs</p> <p>✓ Educate the participants about benefits of physical activity and healthy eating</p> <p>✓ Provide exercise components, provide education components and Counsel pregnant women about nutrition</p> <p>✓ Educate the participants about weight loss</p> <p>✓ Implement intensive marketing, recruitment and retention strategies</p> <p>✓ Implement an intensive quality improvement review plan to assess monthly the program outcomes to drive program modifications as needed</p>	<p>Number of people served, hours of instruction provided and received</p> <p>The direct products of program activities: ✓ Number of classes taught ✓ Number of educational materials distributed ✓ Hours of service delivered ✓ Number of participants served ✓ Number of participants with reduced A1C levels, BMI levels, blood pressure levels and weight loss. ✓ Number of participants who received mammograms and referrals for diagnostic screenings</p>	<p>Benefits as a result of the activities conducted</p> <p>Benefits or changes for participants during or after program activities: <u>Examples must include:</u> ✓ New knowledge ✓ Increased skills ✓ Changed attitudes or values ✓ Modified behavior ✓ Improved Condition (reduced BMI, A1C) ✓ Altered status (no longer pre-diabetic)</p>	<p>What we track and use to measure the benefits to program participants</p> <p>✓ Number and percent of participants that learned how to lose weight (exercise, healthy cooking skills) (pre- and post-test surveys will be administered)</p> <p>✓ Number and percent of participants that lost weight or reduced A1C (a scale will be used to measure participants weight before and after the weight loss class and an A1C test will be administered pre/and post for each cohort but no less than quarterly)</p>

VII. THE OCMH'S SIX PRIORITY AREAS, EXAMPLES OF MANDATORY OUTCOMES AND OUTCOME INDICATORS

The OCMH funds prevention programs in six priority areas, including: 1) cancer prevention education/behavior change and prevention, 2) cardiovascular disease/hypertension prevention education/behavior change, 3) diabetes prevention education/behavior change, 4) infant mortality reduction education/behavior change, 5) substance abuse prevention education/behavior change, and 6) violence prevention education/behavior change. Outcomes objectives are specific statements about the desired changes in the lives of the participants. Outcome objectives identify levels of change in learning, awareness, knowledge, attitudes, skills, and most importantly diagnosis such as from pre-diabetic status to normal A1C levels. Additional outcomes include changes in action, behaviors, practices, and lifestyle decisions. Examples of expected and mandatory outcomes in each of the priority areas can be found in Appendix C.

To standardize the measurement of outcomes across funded OCMH projects, the Commission has based its prevention evaluation strategies on Healthy People 2020 objectives, <https://www.healthypeople.gov/2020/topics-objectives>. As such, the Commission has established a set of 1 to 2 required indicators that must be used, when evaluating Commission funded, prevention programs. Table 1 on the next page shows the required indicators for each priority area.

Projects must set their own numerical targets for the selected indicators through the development of SMART objectives; e.g., “By the end of the (first quarter, year, etc.), 70% of the program participants will obtain a normal A1C level and reduce their BMI by 10%” or “By the end of the (first quarter, year, etc.), 80% of program participants will show a 50% increase in knowledge.” As per the Centers for Disease Control and Prevention, SMART objectives are:

Specific: Concrete, detailed, and well defined so that you know where you are going and what to expect when you arrive

Measurable: Numbers and quantities provide means of measurement and comparison

Achievable: feasible and easy to put into action

Realistic: Considers constraints such as resources, personnel, cost, and time frame

Time-Bound: A time frame helps to set boundaries around the objective Appendix C may be helpful in suggesting appropriate numerical targets for clinical measures.

Other indicators can be selected in addition to the ones required. For example, a diabetes prevention program that includes a nutrition education program could include a satisfaction measure to assess whether participants were satisfied with this component of the program. Appendix C lists a broad range of outcome indicators appropriate to the priority areas. The mandatory indicators must be selected in addition to other outcome indicators that are unique to the project focus and design. For example, a project that aims to increase the use of a particular stress reduction method such as meditation may include a measure that shows a gain in the number of participants who self-report using meditation on a regular basis.

All funded projects must include a lifestyle modification component to include diet, exercise and clinical screenings along with the evaluation of each of these components.

It is required that projects select indicators that document a change in 1) the required clinical measurement such as A1C reduction, body weight reduction, blood pressure reduction , cholesterol level reduction, or other relevant clinical health measurements. In addition, all funded projects must also measure change in increased physical activity as well as knowledge, skills and awareness.

Funded programs are responsible for contracting to for the collection of clinical health measures directly or through partnerships. The quarterly collection of these measures are the ultimate responsibility of the funded project.

TABLE 1. REQUIRED OUTCOMES, OUTCOME INDICATORS AND MEASURES

Type of Prevention Program	Outcome	Outcome Indicator(s)
Cancer Education/Behavior Change	<p>Increased knowledge about cancer prevention strategies</p> <p>Increased number of people receiving preventive screenings for cancer</p>	<ul style="list-style-type: none"> • Percentage of program participants who demonstrate pre- to posttest increase in knowledge about cancer prevention strategies • Number/percentage of individuals receiving project services who receive a cancer screening during the grant period • Percentage of program participants who show a pre- to posttest increase in knowledge about diabetes and behaviors that will reduce the risk of developing cancer • Percentage of program participants who reduce their BMI. • Percentage of program participants who quit smoking. • Percentage of program participants who increase their knowledge levels on the reduction of their risk of developing cancer. • Percentage of program participants who increase their physical activity. <p><i>For this outcome indicator these clinical measures must be used: cancer screening, Weight..</i></p>
Cardiovascular Disease/Hypertension Education/Behavior Change	<p>Increase in knowledge that reduces the risk of developing cardiovascular disease</p> <p>Increase in protective behaviors that reduce the risk of developing cardiovascular disease</p> <p>Improved nutritional decisions</p> <p>Increase healthy cooking skills</p> <p>Increase physical activity</p> <p>Decreased BMI</p> <p>Decreased Cholesterol</p>	<ul style="list-style-type: none"> • Percentage of program participants who reduce their blood pressure. • Percentage of program participants who reduce cholesterol levels • Percentage of program participants who reduce their BMI. • Percentage of program participants who quit smoking. • Percentage of program participants who increase their knowledge levels on the reduction of their risk of developing cardiovascular disease. • Percentage of program participants who increase their physical activity. <p><i>For this outcome indicator these clinical measures must be used: Weight, Blood Pressure, cholesterol screenings.</i></p>

Type of Prevention Program	Outcome	Outcome Indicator(s)
Diabetes Education/ Behavior Change	<p>Increase in knowledge that reduces the risk of developing diabetes</p> <p>Increase in protective behaviors that reduce the risk of developing diabetes</p> <p>Improved nutritional decisions</p> <p>Increase healthy cooking skills</p> <p>Increase physical activity</p> <p>Decreased BMI</p> <p>Decreased A1C levels to normal status</p>	<ul style="list-style-type: none"> • Percentage of program participants who show a pre- to posttest increase in knowledge about diabetes and behaviors that will reduce the risk of developing the disease • Percentage of program participants who reduce their A1C level. • Percentage of program participants who reduce their BMI. • Percentage of program participants who quit smoking. • Percentage of program participants who increase their knowledge levels on the reduction of their risk of developing diabetes. • Percentage of program participants who increase their physical activity. <p><i>We recommend selecting/adapting the Risk Perception Survey for Developing Diabetes (RPSDD) survey developed by the Michigan Diabetes Research and Training Center available at:</i> http://www.med.umich.edu/mdrtc/profs/survey.html.</p> <p><i>The RPSDD was developed for women and would need questions adapted if men are included in the participant population.</i></p> <p>Percentage of program participants who take action to reduce their risk of developing diabetes (for example, changing diet, increasing physical activity, or losing weight)</p> <p><i>For this outcome indicator these clinical measures must be used: Weight, Body Mass Index (BMI) and Hemoglobin A1C. The Hemoglobin A1C measure must be measured for each participant within the cohort on a quarterly basis.</i></p>

<p>Infant Mortality Education/Behavior Change</p>	<p>Increased frequency of prenatal care visits</p> <p>Reduction of Preterm births</p> <p>Increase in normal weights</p> <p>Increased protective factors/reduced risk factors to improve health outcomes for newborn babies through the required 20 pathways of the Certified Pathways Hub Model.</p>	<ul style="list-style-type: none"> Percentage of pregnant women who receive early and adequate prenatal care, as measured by the Kotelchuk Adequacy of Prenatal Care Utilization Index. See the description at: http://health.utah.gov/oph/IBIShelp/kotelchuck.html for additional information. Percentage of pregnant women who give birth to babies at 37 weeks or greater gestational period Percentage of pregnant women who give birth to babies who are not considered low birth weight (i.e., weigh more than 5 lbs., 8 oz.) Percentage of new mothers who report behaviors that reduce the risk of early death for infants (e.g., reducing exposure to secondhand smoke, regular visits to healthcare providers, practice of safe sleeping arrangements, etc.) <p><i>For this outcome indicator these clinical measures must be used: Birth Weight and Birth Term.</i></p>
<p>Substance Abuse Education/Behavior Change</p>	<p>Decreased exposure to alcohol and/or drug using individuals</p> <p>Increased ability to refuse alcohol and/or drugs</p> <p>Increase of individuals who do not use alcohol or other drugs</p> <p>For school age participants the number who have 95% or better school attendance</p>	<ul style="list-style-type: none"> Percentage of program youth participants reporting no substance use at the beginning of the program who continue to report no substance use at the end of the program period Percentage of program youth participants reporting experimental use at the beginning of the program who now to report no substance use at the end of the program period Percentage of program youth participants who avoided alcohol and/or drug environments Percentage of youth participants who attend school at a 95% or better rate. Percentage of program youth participants who declined alcohol/and or drugs when offered to them <p><i>We recommend using the substance use questions from the Youth Risk Behavior Survey (YRBS) for evaluation of this outcome. The survey is available at: http://www.cdc.gov/HealthyYouth/yrbs/index.htm</i></p>
<p>Violence Education/Behavior Change</p>	<p>Increased awareness of types of violent acts (verbal, physical, sexual, and/or electronic)</p>	<ul style="list-style-type: none"> Percentage of program participants who understand what violence is and their ability to identify violent acts Percentage of youth participants who attend school at a 95% or better rate.

	<p>Increased involvement in constructive activities/social events</p> <p>Increased ability to avoid peers who have poor conduct.</p> <p>For school age participants the number who have 95% or better school attendance</p> <p>Increased ability to avoid gang involvement or gang related activity</p> <p>Decreased use of social media to escalate conflicts</p>	<ul style="list-style-type: none"> • Percentage of program participants who report involvement in positive skill/social development activities • Percentage of program participants who report avoiding peers who have poor conduct. • Percentage of program participants who report avoidance of gang involvement or related activity. • Percentage of program participants who avoid potential incidents of physical altercations, bullying, harassing and/or electronic aggression <p><i>We recommend selecting measurement instruments from the CDC publication <u>Measuring Violence-Related Attitudes, Behaviors, and Influences Among Youths: A Compendium of Assessment Tools</u> available at http://www.cdc.gov/ncipc/pub-res/measure.htm</i></p> <p><i>The violence questions from the YRBS may also be used for assessment of outcomes. These are available at: http://www.cdc.gov/HealthyYouth/yrbs/index.htm</i></p>
--	--	--

VIII. EVALUATION APPROACHES

A. EVALUATION DESIGNS

Evaluation Designs

Evaluation “design” is a term used by evaluators to describe a scientific strategy to systematically measure program effect and to attribute that effect to a program’s intervention. A large number of evaluation designs are available for use by grantees.³ Evaluators recognize that some designs are much “stronger” than others, however. Strength is measured by the design’s ability to enable the evaluator to conclude that a program effect is due to the program and not to something else. The “something else” refers to all the other possible explanations that typically exist for a program’s effect, usually referred to as the “threats to the internal validity” of a program.⁴ All OCMH grantees are required to identify and discuss an evaluation design in their evaluation proposal.

While the “gold standard” in evaluation design in general is the randomized comparison group design, the OCMH understands that this design is unlikely to have applicability in a local community setting. Therefore, the OCMH recommends at a minimum that a one-group pre-test/post-test design be used with outcome measures collected from all program participants using valid and reliable instruments. The pre-test/post-test design can be strengthened considerably by adding a similar comparison group for the data collection who will not be involved during the programming phase. No matter what design you choose, however, baseline (pre-implementation) data must be collected on program participants before the start of the program to establish a benchmark to compare change against. You will then be expected to collect the same information from participants after the program is complete.

Choosing Your Evaluation Design

Evaluation designs must be selected to fit a particular program’s situation. There is no “one size fits all” evaluation design in most cases. Grantees are **strongly encouraged** to discuss their evaluation design with their evaluation consultant at the proposal development stage. It is easier to “fit” an evaluation design to a program at the program planning stage than it is after implementation. Your evaluation consultant will guide you on the proper selection of an appropriate evaluation design.

B. INSTRUMENTATION, RELIABILITY AND VALIDITY

An instrument is what we use to obtain information on the knowledge, skills, attitudes and behavior of participants in a health promotion project. The term instrument is often used interchangeably with the terms tool, survey, measure, measurement, and questionnaire. It may be also referred to as a scale, or a cluster of items that measure a single concept; for example, an

³ You may have heard of some of these designs referred to by such terms as the one group pre-test/posttest design, the randomized comparison group design, or simply as a quasi-experimental design.

⁴ See glossary of terms for a definition of internal validity.

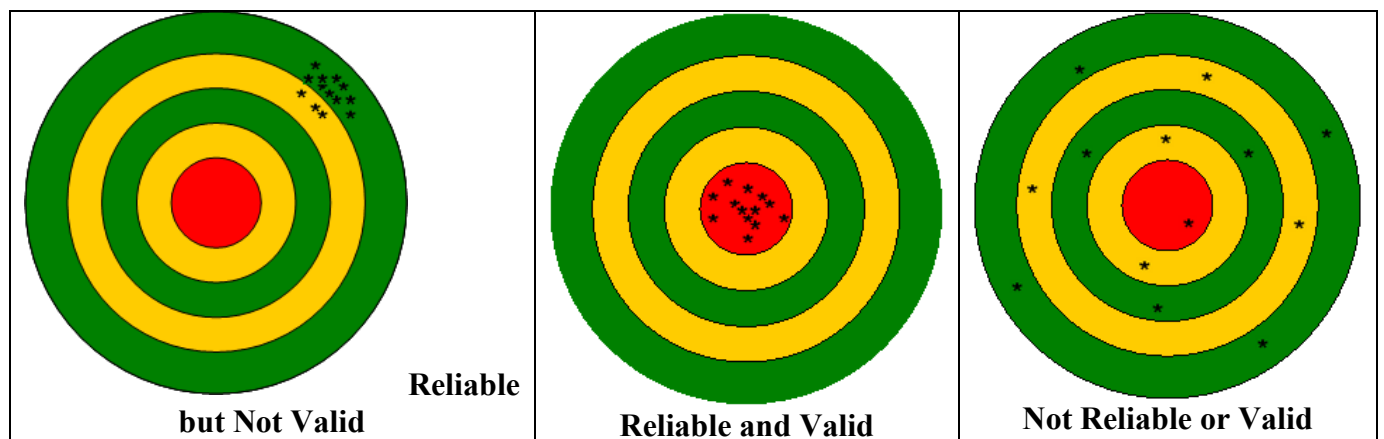
instrument measuring diabetes-related behavior might have a scale that measures “knowledge about diabetes” and another that measures “actions taken to manage diabetes.”

One way in which the quality of an instrument can be verified is by knowing that it is reliable and valid.

Reliability. Reliability refers to the consistency of an instrument to yield the same results when the instrument is used to measure the same thing time and time again. An instrument is said to be reliable when it consistently measures each time it is used. A reliable instrument is like a reliable watch—its accurate keeping of time is not affected by the race, weight, height, or income of the person wearing it nor by the month of the year, day of the week, or temperature of the room in which it is worn. If an item on a survey is unclear, vague, or subject to various interpretations, then it might be understood differently by the person responding and the responses to the survey questions would therefore not be reliable.

Validity. Validity refers to the degree to which an instrument measures what it was designed to measure. For example, if you are trying to measure the impact of a diabetes education program focusing on healthy nutritional choices but your instrument only asks questions about attitudes toward food, it would not be a valid measure of the effects of your health education intervention.

One way to understand the concepts of reliability and validity and their importance to evaluation research is to envision an archer and a target.⁵ In this context, an evaluation instrument is like an arrow shot at a target and the bulls’ eye on the target is what we are trying to measure. Over repeated attempts to connect with the target we will be able to observe the level of consistency (reliability) and accuracy (validity) of our “arrow.” The center panel below reflects the experience of using an instrument that is both reliable and valid. To the left we can see the effects of using an instrument that is reliable but is off target for what we are seeking to learn. To the right, our instrument is both unreliable and not valid, resulting in scattered shots that do not hit the bulls’ eye.



⁵ Graphic adapted from <http://www.georgetown.edu/departments/psychology/researchmethods/researchanddesign/validityandreliability.htm>

The reliability and validity of an instrument has a direct impact on the confidence that can be placed in the evaluation findings. This, in turn, significantly impacts the usefulness of the evaluation findings in planning future programs and in anticipating future outcomes.

Many of the measures of the effects of a health promotion are clinical indicators (e.g., blood pressure readings) or behaviors directly observed (e.g., participants exercising in a fitness program). While they do of course require documentation, they obviously do not involve instrumentation that needs validity and reliability established. However, many intermediate outcome indicators (e.g., attitudes toward participating in an exercise program or self-reported behaviors to lower blood pressure) do require reliable and valid measures to assess changes in participant attitudes or behaviors. **Whenever possible, it is highly desirable to use standardized instruments that have already been established as having sound reliability and validity.** Use of these instruments has the advantage of yielding evaluation findings that can be presented with greater confidence and credibility. An excellent source of information on behavioral or clinical measurement instruments is the Health and Psychosocial Instruments (HaPI) database of the Behavioral Measurement Database Services. HaPI identifies, catalogs, and disseminates information on health and behavioral measures used in medicine, nursing, public health, psychology, social work, and related fields. Accessing HaPI requires a subscription. However, it is likely that your evaluator, if he or she is associated with an academic institution, will have access to this or similar services.

The lack of availability of instruments in some areas may necessitate the development of a new instrument. In such cases it is important for the agency and the evaluator to work together to assess the new instrument's reliability and validity (sometimes referred to as psychometric properties). To learn more about this process, readers are referred to: DeVellis, R. F. (2003), *Scale Development: Theory and Applications*.

C. DATA COLLECTION METHODS⁶

A key aspect of evaluation involves the selection of appropriate data collection methods. Table 2 presents information about six types of data collection methods that may be relevant in specific program environments. The table shows each method's primary purpose and the corresponding advantages and challenges.

Using Single versus Multiple Methods

No one data collection method is ideal for every situation. For this reason, it is preferable to use multiple methods whenever possible. Using multiple methods to assess the same outcomes (e.g., using surveys and document reviews to assess program management) provides a richer, more detailed picture. It also illuminates inconsistencies between methods and reduces the chance of bias caused by a particular method.

Quantitative and Qualitative Evaluation Methods

Quantitative evaluation methods are required in most evaluation situations. Quantitative

⁶ Section extracted from: Bouffard, S., and Little, P. M. D. (2004).

evaluation attempts to use systematic methods to quantify or measure program outputs and outcomes. Thus, a quantitative evaluation of a typical program will count the number of program sessions offered to program participants, count the number of clients who attended the program sessions or ask participants for their evaluation of service delivery on a standardized satisfaction survey at the completion of the program. Similarly, a quantitative evaluation attempts to measure outcomes such as changes in the skills, attitudes or behaviors of program participants using reliable and valid instruments. Quantitative evaluation can be strengthened considerably by also adding qualitative approaches.

In qualitative evaluation the meaning of any situation is perceived from the participant's viewpoint, and the researcher is a mere observer. Qualitative evaluation attempts to be holistic in its orientation to construct and interpret reality from multiple sources. Qualitative designs are dynamic and based on inductive reasoning. Some commonly used methods that employ qualitative designs for conducting health program evaluations include: ethnographies, phenomenologies, interviews, case studies and participatory research studies (Sharma, 2004). The sampling techniques that are often used for qualitative evaluation include purposive and convenience sampling. Interviews, focus groups and observations are common methods for collecting data. Content analysis is then used to analyze the results from the data collected. Issues of trustworthiness (adherence to procedures and rigor), credibility (consensus), coherence (extent to which the final evaluation makes sense), transferability (ideas about similarity of evaluation in other settings), dependability (similarity with other evaluations), and confirmability (extent to which neutrality of evaluation can be established) are important considerations. For an example of participatory research see a study conducted by Sharma and Deepak (2001). The data collection methods mentioned above are summarized in Table 2

Selecting Data Sources

Equally important in selecting a data collection method is selecting a data source or information provider. Data can be collected from youth, families, staff, funders, educators, and other stakeholders. Certain data sources lend themselves more easily to certain data collection methods; for example, to assess parent satisfaction, surveys are popular because they can be mailed to hard-to-reach parents. As with data collection methods, it is recommended that evaluations use several data sources. However, programs may already have some of the evaluation information they need. Conducting an inventory of what programs already have and what is already available in the community can reduce the data burden as well as data collection costs.

Selecting a Sample of Individuals

The choice of data collection methods is also affected by the sample to be studied. Some methods are well suited to collecting data from all participants (e.g., surveys), while others are better suited for a smaller group that represents the diversity of all participants (e.g., focus groups). The choice of sample is in turn affected by the size of the program to be evaluated. In general, including all participants produces the most reliable results but may limit the type and amount of data collected because of the cost implications of doing so.

Collecting Data Before and After Program Participation and Implementation

Using the same data collection method to gather information before the start of the program and after its completion (also known as a *pre-/post-test design*) provides the opportunity to determine whether some characteristic changed during the course of the program. This can *suggest* that the program played a role in effecting the change; however, unless a program uses random assignment as part of its pre/posttest design, this method cannot establish that the program *caused* the change, because other unmeasured factors may have been responsible.

Cost Considerations

Selecting data collection methods has cost implications. While conducting observations and reviewing program documents can be done with little additional funding, designing and administering a survey and analyzing its results may require the assistance of an outside evaluator. At a minimum, it will require additional staff time for training. Similarly, using standardized tests and assessments will likely necessitate external expertise to analyze results. An additional cost consideration is the use of management information systems (MIS) that can be used to record and store data collected from numerous methods. Like most other components of evaluation, it is important to start small and build data collection tools over time, as appropriate to the program's evolving evaluation needs.

Table 2: Data Collection Methods

Method	Purposes	Advantages	Challenges
Clinical Screens	Collected by trained medical or non-medical professionals to gather specific clinical data from participants. Administered in person using medical equipment such as scales, blood tests, stethoscopes, etc.	<ul style="list-style-type: none"> <input type="checkbox"/> Effective in showing immediate results and program impact <input type="checkbox"/> Can be administered to large groups of individuals <input type="checkbox"/> Effective for assessing program results <input type="checkbox"/> Provides rich data that can highlight areas for program improvement <input type="checkbox"/> More valid and reliable than perceptions, self-report or opinions <input type="checkbox"/> Can be used to project cost savings and return on investment 	<ul style="list-style-type: none"> <input type="checkbox"/> May require medical personnel <input type="checkbox"/> Will require budgeting to cover costs <input type="checkbox"/> Will require intense education of program participants on purpose of clinical screening
Tests and Assessments	Developed or used specifically for the program evaluation to quantify characteristics of the program, participants or outcomes. Examples include physical exams and psychological tests. May be standardized or created by program evaluators for the specific program.	<ul style="list-style-type: none"> <input type="checkbox"/> Often more valid and reliable than perceptions or opinions. <input type="checkbox"/> Comparing pre-/post-test scores is a good method for assessing change. 	<ul style="list-style-type: none"> <input type="checkbox"/> Can be costly and time intensive. <input type="checkbox"/> May require scoring by an external source.
Surveys and Questionnaires	Collected by evaluators to gather specific information from participants, families, staff and administrators, community members, and other stakeholders. Usually administered on paper, in a structured or semi-structured format. Respondents often choose from among a set of <i>forced-choice</i> , or provided, responses. Can be administered in person, by mail, over the phone, or via email/Internet.	<ul style="list-style-type: none"> <input type="checkbox"/> Less time consuming and expensive to administer. <input type="checkbox"/> Can be administered to large groups of individuals. <input type="checkbox"/> Effective for assessing program satisfaction. 	<ul style="list-style-type: none"> <input type="checkbox"/> Data entry and analysis can be time consuming. <input type="checkbox"/> May be difficult to receive completed surveys from stakeholders. Incentives can be offered to increase response rate.
Record or Document Reviews	Analyze existing program records and other documents not gathered or developed specifically for the evaluation. Examples include recruitment and attendance records, budget, staff records and annual reports. Particularly useful for documenting implementation.	<ul style="list-style-type: none"> <input type="checkbox"/> Records are tailored to programs. <input type="checkbox"/> Saves on evaluation time/cost. <input type="checkbox"/> May elicit a high degree of accuracy if records kept. 	<ul style="list-style-type: none"> <input type="checkbox"/> May not be available or applicable for some indicators. <input type="checkbox"/> May be incomplete due to lack of priority on documentation.
Interviews and Focus Groups	Most often used to gather detailed, qualitative descriptions of how programs operate and how stakeholders perceive them. Interviews are conducted one-on-one, while focus groups are conducted in small	<ul style="list-style-type: none"> <input type="checkbox"/> Provide rich data that paint a broad picture. 	<ul style="list-style-type: none"> <input type="checkbox"/> Interviews and focus groups may intimidate some participants.

	groups. Usually conducted with targeted samples of stakeholders. Can be conducted in person or by phone. Questions are generally open-ended and responses documented in detailed notes or transcription.	<input type="checkbox"/> May highlight information useful for interpreting quantitative data. <input type="checkbox"/> Small focus groups may increase the comfort level of participants.	<input type="checkbox"/> Documentation and analysis can be time consuming and may require the help of someone versed in qualitative analysis.
Observations	Generally an unobtrusive method for gathering information about how the program operates. Usually conducted by external evaluators and are often used to supplement information gathered through other methods. Can be highly structured or unstructured. More reliable if conducted over time to lessen the chance that observations are atypical.	<input type="checkbox"/> Provide highly detailed information from an external perspective on what occurs. <input type="checkbox"/> Trained observers may provide less biased descriptions.	<input type="checkbox"/> Labor intensive and expensive. <input type="checkbox"/> Observers must be trained and be consistent. <input type="checkbox"/> Observations may not represent practices over time.
Secondary Sources and Data Reviews	Use existing documents or data that were originally collected for other purposes, but which are useful for the evaluation. Examples include achievement data, standardized test scores, court records and community demographic data.	<input type="checkbox"/> May be less biased than perceptions or opinions. <input type="checkbox"/> Can save administrative time and costs.	<input type="checkbox"/> Obtaining records often requires special permission. <input type="checkbox"/> Some national datasets charge access fees.

Summarized from: Bouffard, S., and Little, P. M. D. (2004). Detangling Data Collection: Methods for Gathering Data. Harvard Family Research Project, Harvard Graduate School of Education.

IX. REFERENCES

- Airhihenbuwa, C. O. (1995). *Health and culture: beyond the western paradigm*. Thousand Oaks, CA: Sage.
- Arnold, C. L., Davis, T. C., Berkel, H. J., Jackson, R. H., & Nandy, I. (2001). Smoking status, reading level, and knowledge of tobacco effects among low-income pregnant women. *Preventive Medicine*, 32, 313-320.
- Bennett, M. J. (1993). Towards ethnorelativism: A developmental model of intercultural sensitivity. In R. M. Paige (Ed.), *Education for the intercultural experience*. Yarmouth, ME: Intercultural Press.
- Berkman N.D., DeWalt D.A., Pignone M.P., Sheridan S.L., Lohr K.N., Lux L., et al. (2004). *Literacy and health outcomes*. Evidence Report/Technology Assessment No. 87 (Prepared by RTI International–University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016). AHRQ Publication No. 04-E007-2. Rockville, MD: Agency for Healthcare Research and Quality. Retrieved February 3, 2006 from: <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hstat1a.chapter.32213>
- Bouffard, S., & Little, P. M. D. (2004). *Detangling data collection: methods for gathering data*. Harvard Family Research Project, Harvard Graduate School of Education. Retrieved February 3, 2006 from <http://www.gse.harvard.edu/hfrp/projects/afterschool/resources/snapshot5.html>
- Bresolin, L. (1999). Health literacy. *The Journal of the American Medical Association*, 281, 552-557.
- Davis, T. C., Frederickson, D. D., Arnold, C., Murphy, P. W., Herbst, M., & Bocchini, J. A. (1998). A polio immunization pamphlet with increased appeal and simplified language does not improve comprehension to an acceptable level. *Patient Education Counsel*, 33, 5-37.
- Deason, L. M. & Wallace, S. L. (2005). Does a shared culture between nursing home staff and residents influence quality of life? *International Journal of Diversity in Organisations, Communities, and Nations*, 4, 235-242.
- Endo, T., Joh, T., & Yu, H. C. (2003). *Voices from the field: health and evaluation leaders on multicultural evaluation*. Woodland Hills, CA: The California Endowment.
- Govindarajan, G., Whaley-Connell, A., Mugo, M., Stump, C., & Sowers, J. R. (2005). The cardiometabolic syndrome as a cardiovascular risk factor. *American Journal of Medical Science*, 330, 311-318.
- Institute of Medicine. (2002). *Unequal treatment: confronting racial and ethnic disparities in health care*. Washington, DC: National Academy Press.

- Institute of Medicine. (2004). *Health literacy: a prescription to end confusion*. Washington, DC: The National Academy Press.
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards*. Thousand Oaks, CA: Sage.
- Lenburg, C., Lipson, J. G., Demi, A. S., Blaney, D. R., Stern, P. N., Schultz, P. R., et al. (1995). *Promoting cultural competence in and through nursing education: a critical review and comprehensive plan for action*. Washington, DC: American Academy of Nursing.
- Michielutte, R., Bahnson, J., Dignan, M. B., & Schroeder, E. M. (1992). The use of illustrations and narrative text style to improve readability of a health education brochure. *Journal of Cancer Education*, 7, 251-260.
- Murphy, P. W., Chesson, A.L., Walker, L., Arnold, C. L., & Chesson, L. M. (2000). Comparing the effectiveness of video and written material for improving knowledge among sleep disorders clinic patients with limited literacy skills. *Southern Medical Journal*, 93, 297-304.
- National Library of Medicine. (2000). *Current bibliographies in medicine: health literacy*. Bethesda, MD: National Institutes of Health.
- Royce, D. (1992). *Program evaluation*. Chicago: Nelson-Hall.
- Sharma, M., & Deepak, S. (2001). A participatory evaluation of community-based rehabilitation program in North Central Vietnam. *Disability and Rehabilitation*, 23, 352-358.
- Sharma, M. (2004). Qualitative research methods. In K. E. Berg & R. W. Latin, *Essentials of modern research methods in health, physical education, and recreation* (2nd ed., pp. 213-224). Philadelphia: Lippincott.
- Tripp-Reimer, T. (1997). Ethnicity, aging and chronic illness. In L. Swanson and T. Tripp-Reimer (Eds.), *Chronic illness and the older adult* (pp. 112-135). New York, NY: Springer.
- United Way of America. (1996). *Measuring program outcomes: a practical approach*. Alexandria, VA: United Way of America.
- Williams, M.V., Baker, D.W., Honig, E.G., Lee, T.M., & Nowlan, A. (1998). Inadequate literacy is a barrier to asthma knowledge and self-care. *Chest*, 114, 1008-1015.
- Williams, M. V., Baker, D. W., Parker, R. M., & Nurss, J. R. (1998). Relationship of functional health literacy to patients' knowledge of their chronic disease. A study of patients with hypertension and diabetes. *Archives of Internal Medicine*, 158, 166-172.

APPENDICES

Appendix A. Bennett's Model of Cultural Competence

There are many developmental models of cultural competence in the field. One commonly referenced model was created by Milton Bennett (1993) and consists of 6 stages of moving from ethno-centrism to ethno-relativism. The 6 stages of the Bennett model include:

1. Denial: an individual denies that cultural differences exist.
2. Defense: an individual acknowledges the existence of certain cultural differences, but because those differences are threatening to her/his own reality and sense of self, the individual constructs defenses against those differences.
3. Minimization: an individual acknowledges cultural differences, but trivializes them, believing that human similarities far outweigh the differences.
4. Acceptance: an individual recognizes and values cultural differences without evaluating those differences as positive or negative.
5. Adaptation: an individual develops and improves skills for interacting and communicating with people of other cultures.
6. Integration: an individual not only values a variety of cultures, but are constantly defining their own identity and evaluating behavior and values in contrast to and in concert with a multitude of cultures. Rising above the limitations of living in one cultural context, these individuals integrate aspects of their own original cultural perspective with those of other cultures.

Appendix B. Culturally Competent Methodologies

Mixed Method Approaches. Mixed method approaches are at the heart of most descriptions of multicultural evaluation strategies. Qualitative methods—such as interviews, focus groups, ethnographic studies or observations—are described as critical for capturing a rich level of data on the cultural context of the communities under study. As much as surveys can quantitatively document differences in outcomes, interviews or focus groups can be useful in gathering more nuanced information on *contributing factors* to those differences. This is key especially within diverse populations where issues such as cultural misunderstanding, socio-economic barriers (i.e., transportation, child care, etc.), or even institutionalized racism may otherwise go undetected. Observation and ethnography can be important for capturing the nuances of cross-cultural nonverbal communication (i.e., between teachers and students) to get closer to the understanding of truth. Further, qualitative approaches such as interviews or focus groups were described as simply more appropriate for gathering accurate data on specific populations that may have built a resistance to written surveys. While respondents highlight qualitative data as a method necessary for culturally competent evaluations, none do so to the exclusion of quantitative approaches. Rather, most health and evaluation leaders with whom we spoke conclude that using mixed methods—using both quantitative and qualitative data sources in culturally competent evaluation—helps reign in cultural bias, leads to more thorough research and gives evaluators more evidence to counteract accusations of non-validity to any one research method. Mixed method approaches were described by some as especially critical within the health field, where data such as rates of access and utilization of services, health insurance coverage, disease incidence, etc., drive decision-making and funding at the federal level.

[Qualitative methods] afford the evaluator an opportunity to get data that is more rich than a quick and dirty survey...to tease out what would be viable interventions to problems that [target populations] face. We have not paid enough attention to mining the wisdom and insight of these folks. —Ricardo Millett, funder

Multiculturalism to me means the way you put an issue, a problem, a study, a questionnaire in context...and then tools are appropriately applied...knowing when and how to use them is going to be informed by your understanding of the context, how you develop questions, translate the questions, interpret interviews, etc. can all be shaped by the way we understand and contextualize a particular given problem.—Ricardo Millett, funder

Triangulation. Triangulation involves using independent or different sources of measurement that corroborate the same findings and is also offered as a strategy. Triangulation offers similar benefits as using mixed methods in an evaluation study, and has an added benefit of multiple sources of data supporting the same findings to evaluation questions. For example, in evaluating an HIV-prevention program, one evaluator used quantitative HIV/AIDS infection rates data and then complemented them with ethnographic profiles of those going through the prevention program. Another evaluator described how they triangulate data obtained from reports, interviews, written surveys and site visits. They use these four data sources to look for convergence and divergence among the sources.

Reflexive staffing. Several health leaders and evaluators mention recruiting evaluators who reflect the communities studied as another important culturally competent evaluation strategy. Ensuring that evaluation teams are composed of professionals who reflect the community is important for a number of reasons. First, diverse staff often possesses knowledge about the culture, traditions and behavioral patterns of their own ethnic group because they were raised in a

manner similar to other group members. Second, sharing a particular racial or ethnic background helps evaluators to be more sensitive to particular issues because there is a likelihood that they themselves have faced similar situations (e.g., racial discrimination and heightened awareness of power differentials that exist between majority and minority communities). This awareness of cultural differences and nuances are likely to shape how they develop or adapt outcome indicators and valid cross-cultural measurements of program “success” or “failure.” Third, appearance matters. When community members see an evaluator from their ethnic or racial background, they are more likely to think that the person is someone with whom they can identify and trust, and that the evaluation will respond with greater sensitivity to their concerns.

It isn't easy for nonminority evaluators to understand issues of inclusion and power...More often than not, non-minority evaluators are more geared toward the bottom line, getting efficient and effective answers, sacrificing the sloppy and often tough work of understanding context and would more favor taking the shortcut.

While it is clear that hiring diverse evaluation staff can improve the quality of evaluations, at the same time, the evaluation and health leaders urge the adoption of a more sophisticated definition of culturally competent evaluators that look beyond race and ethnicity. As discussed earlier, culturally competent evaluators and culturally appropriate practices are not just a function of race and ethnicity, but are also influenced by a multitude of other factors such as cultural heritage, geographic community and socioeconomic status.

Attention to reflexive staffing is an important initial consideration, but capitalizing on diversity means that all individuals who work with communities of difference need to be prepared to establish bonds of trust across gaps of culture and race.

Engagement of Community Members. A few health leaders and evaluators describe engagement of community members *themselves* as a strategy for gaining cultural insight on a community and ensuring that the design and implementation of an evaluation honors a community’s cultural context. Engaging community members as co-researchers empowers community-based groups and builds their ability to conduct self-evaluation. One evaluation leader gave an example of how inclusion of communities in the evaluation design helped to inform their context. When asking different home care workers across four ethnic groups their conception of mental illness, the study found that each group had a different concept of what it meant to be mentally ill. Because of the four different concepts, defining the outcomes, then, became more complicated and time consuming, yet this added measure led to a better evaluation. Some evaluators describe an approach of always making sure a “cultural translator” is a part of their evaluation team so they are sure they are capturing nuances of a culture that might be different than theirs. A cultural translator is either an evaluator him or herself, a research associate, or a beneficiary stakeholder. For example, one African American evaluator described doing some evaluation with a Native American community, where he employed a Native American research associate on his team. This team member was able to describe the significance of a community baking an earth cake and how this process helped to describe the community mental health of the Navajo tribe. This evaluator also described how another cultural translator trained in survey design was able to point out to him that a Likert scale (a type of survey response scale) does not particularly work well with the Hmong population, who generally see a program as “good”, or “bad” and nothing in between.

Appendix C. Outcome Indicators for Health Promotion/Disease Prevention Programs

Note: many of the short-term, intermediate, and long term outcomes were taken from *Healthy People 2020*, the ten year health objectives established by the U.S. Department of Health and Human Services in the year 2010. Targeted to populations with Health Disparities: Use surveillance data to identify vulnerable populations; choose evidence-based programs, policies, and practices; and choose culturally and linguistically appropriate services.

CANCER PREVENTION			
Outputs (process outcomes)	Initial Outcomes	Intermediate outcomes	Long-term outcomes*
<ul style="list-style-type: none"> ▪ Number of participants receiving educational intervention. ▪ Number of health care providers receiving intervention to increase screening rates. ▪ Number of hours of educational intervention provided 	<p>Outcome Indicators:</p> <ul style="list-style-type: none"> ▪ participants showing improvement in knowledge, attitude, and/or skills scores related to: <ul style="list-style-type: none"> ○ cancer screening, ○ healthy dietary changes, ○ physical activity or ○ tobacco cessation. ▪ Number of women counseled by their providers regarding pep tests 	<p>Outcome Indicators:</p> <ul style="list-style-type: none"> ▪ People who use skin cancer prevention measures ▪ Proportion of adolescents (grade 9-12) and adults (over 18) who use artificial sources of UV lightening for tanning ▪ Proportion of people who experience sun burns ▪ Women over 18 who have ever: <ul style="list-style-type: none"> ○ received a Pap test ○ received a Pap test within the past 3 years ▪ Women aged 40 years and older who have received a mammogram within the previous 2 years. ▪ Adults over 50 years who: <ul style="list-style-type: none"> ○ received a fecal occult blood test (FOBT) within the preceding 2 years for colorectal cancer screening ○ received a sigmoidoscopy for colorectal cancer screening ▪ Proportion of adults who received colorectal cancer screening ▪ Persons aged 2 years and older who consume: <ul style="list-style-type: none"> ○ at least two daily servings of fruit ○ at least three daily servings of vegetables, 	<p>Outcome Indicators:</p> <ul style="list-style-type: none"> ▪ Overall reduction in cancer death rates. ▪ Overall reduction in organ-specific death rates: <ul style="list-style-type: none"> ○ Lungs ○ Female breast ○ Uterine cervix ○ Colorectal ○ Oropharyngeal ○ Prostate ○ Melanoma ○ Invasive rectal ○ Invasive uterine cervical ▪ Cancer survivors living 5 years or longer after diagnosis.

* Most long term outcomes will be beyond the ability of health promotion programs to effect in a limited project time. They are included in the table to illustrate what short term and intermediate outcomes are meant to eventually achieve.

CANCER PREVENTION			
Outputs (process outcomes)	Initial Outcomes	Intermediate outcomes	Long-term outcomes*
		<ul style="list-style-type: none"> with at least one-third being dark green or orange vegetables <ul style="list-style-type: none"> ○ < 10 % of calories from saturated fat. ○ < 30 % of calories from total fat. ▪ Adults who, in the past 12 months, report having had an examination to detect oral and pharyngeal cancers. ▪ Cigarette/spit tobacco/cigar use by adults ▪ Cigarette/spit tobacco/cigar use by adolescents in the past month. ▪ Smoking cessation attempts by adult smokers. ▪ Tobacco use cessation attempts by adolescent smokers. ▪ Physicians and dentists who counsel their at-risk patients about tobacco use cessation, physical activity, and cancer screening. ▪ Oral and pharyngeal cancers detected at the earliest stage in the targeted populations. ▪ Proportion of men who have discussed with their health care provider about having prostate specific antigen (PSA) test to screen for prostate cancer 	

CARDIOVASCULAR DISEASE PREVENTION			
Outputs (process outcomes)	Initial Outcomes	Intermediate outcomes	Long-term outcomes
<ul style="list-style-type: none"> ▪ Number of participants receiving educational intervention. ▪ Number of participants receiving physical activity intervention. ▪ Number of educational/other interventions offered. 	<p>Outcome Indicators:</p> <ul style="list-style-type: none"> ▪ adolescents who disapprove smoking. ▪ participants showing improvement in knowledge/attitude/skill scores related to: 	<p>Outcome Indicators:</p> <ul style="list-style-type: none"> ○ Increased tobacco cessation. ○ Reduced Cholesterol levels ○ Reduced BMI ○ Reduced Blood Pressure <ul style="list-style-type: none"> ▪ Adults who have had their baseline blood cholesterol checked at the beginning of the 	<p>Outcome Indicators:</p> <ul style="list-style-type: none"> ▪ increased overall cardiovascular health ▪ decreased coronary heart disease death rates. ▪ decreased stroke death rates ▪ decreased mean total blood

CARDIOVASCULAR DISEASE PREVENTION

Outputs (process outcomes)	Initial Outcomes	Intermediate outcomes	Long-term outcomes
<ul style="list-style-type: none"> ▪ Number of activities participated in. ▪ Number of hours of direct contact with program participants. 	<ul style="list-style-type: none"> ○ early signs of heart disease and heart attack, ○ stress reduction, ○ Increased knowledge of healthy dietary changes ○ Increased physical activity ○ 	<p>program.</p> <ul style="list-style-type: none"> ▪ Adults who improve blood pressure levels within the program cohort. ▪ Older adults who have participated during the program year in health education promotion activity related to cardiovascular disease. ▪ Persons aged 2 years and older who consume: <ul style="list-style-type: none"> ○ at least two daily servings of fruit. ○ at least three daily servings of vegetables, with at least one-third being dark green or orange vegetables. ○ <10 % of calories from saturated fat. ○ <30 % of calories from total fat. ▪ Adults who: <ul style="list-style-type: none"> ○ engage regularly, preferably daily, in moderate physical activity for at least 30 minutes per day. ○ engage in vigorous physical activity that promotes the development and maintenance of cardiorespiratory fitness 3 or more days per week for 20 or more minutes per occasion. ▪ Adolescents who: <ul style="list-style-type: none"> ○ engage in moderate physical activity for at least 30 minutes on 5 or more of the previous 7 days. ○ engage in vigorous physical activity that promotes cardiorespiratory fitness 3 or more days per week for 20 or more minutes per occasion. ○ adolescents who view television 2 or fewer hours on a school day. ▪ Adolescents and adults who make trips by: <ul style="list-style-type: none"> ○ Walking. ○ Bicycling. ▪ Proportion of adults aged 20 years and older who are aware of, and respond to, early warning 	<p>cholesterol levels among adults.</p> <p>Reduced number of:</p> <ul style="list-style-type: none"> ▪ adults with high total blood cholesterol levels. ▪ Adults who are at a healthy weight. ▪ Adults who are obese. ▪ children and adolescents who are obese. ▪ nonsmokers exposed to environmental tobacco smoke as measured by serum cotinine levels. ▪ LDL levels at or below recommended levels among adults (with heart attacks or strokes)

CARDIOVASCULAR DISEASE PREVENTION			
Outputs (process outcomes)	Initial Outcomes	Intermediate outcomes	Long-term outcomes
		symptoms and signs of a heart attack <ul style="list-style-type: none"> ▪ Proportion of adults aged 20 years and older who are aware of, and respond to, early symptoms and signs of stroke 	

DIABETES PREVENTION			
Outputs (process outcomes)	Initial Outcomes	Intermediate outcomes	Long-term outcomes
<ul style="list-style-type: none"> ▪ Number of participants receiving educational intervention. ▪ Number of participants counseled about diet. ▪ Number of educational interventions offered. ▪ Number of hours of direct contact with program participants. 	Outcome Indicators: <ul style="list-style-type: none"> ▪ participants showing improvement in knowledge/attitude/skills scores related to: <ul style="list-style-type: none"> ○ early signs of diabetes detection, ○ healthy dietary changes, or ○ Self-efficacy ○ Social support 	Outcome Indicators: <ul style="list-style-type: none"> ○ Increased physical activity. ○ Moderate and sustained weight loss ○ Reduced A1c to normal levels ○ Reduced BMI ○ Referrals to physician office by patients at-risk patients ○ Proportion of patients who are pre-diabetic (A1c value - 5.7 to 6.4 percent) ○ Proportion of Adults who have A1c in the normal range ○ Adults who: <ul style="list-style-type: none"> ○ engage regularly, preferably daily, in moderate physical activity for at least 30 minutes per day. ○ engage in vigorous physical activity that promotes the development and maintenance of cardiorespiratory fitness 3 or more days per week for 20 or more minutes per occasion. ○ Adolescents who: <ul style="list-style-type: none"> ○ engage in moderate physical activity for at least 30 minutes on 5 or more of the previous 7 days. ○ engage in vigorous physical activity that promotes cardiorespiratory fitness 3 or 	Outcome Indicators: <ul style="list-style-type: none"> ▪ Reduced Incidence of Type 2 Diabetes ▪ Increased number of adults who are at a healthy weight. ▪ Reduced numbers of obese children, adolescents and adults who are obese. ▪ Increased control of LDL cholesterol values among adults that are pre-diabetic.

DIABETES PREVENTION			
Outputs (process outcomes)	Initial Outcomes	Intermediate outcomes	Long-term outcomes
		<p>more days per week for 20 or more minutes per occasion.</p> <ul style="list-style-type: none"> ○ Persons who have an increase knowledge, skill level and report behavior change about: <ul style="list-style-type: none"> ○ physical activity or exercise. ○ diet and nutrition. ○ smoking cessation. 	

INFANT MORTALITY			
Outputs (process outcomes)	Initial Outcomes	Intermediate outcomes	Long-term outcomes
<ul style="list-style-type: none"> ▪ Number of participants receiving educational intervention. ▪ Number of participants counseled about diet in pregnancy. ▪ Number/hours of educational interventions offered. ▪ Number of hours of direct contact with program participants. 	<p>Outcome Indicators:</p> <ul style="list-style-type: none"> ▪ participants showing improvement in knowledge/attitude/skills scores related to infant mortality reduction behaviors and activities. ▪ Increased Prenatal visits ▪ Increased smoking cessation 	<p>Outcome Indicators:</p> <ul style="list-style-type: none"> ▪ Infants who are put down to sleep on their backs. ▪ Abstinence from alcohol, cigarettes, and illicit drugs among pregnant women. ▪ Reduced number of preterm births ▪ Reduced number of Low birth and very low birth weight babies ▪ Increase in Normal birth weight babies ▪ Increase in post-partum visit completion ▪ Mothers who: <ul style="list-style-type: none"> ○ breastfeed their babies in early post-partum period. ○ breastfeed their babies at 6 months. ○ breastfeed their babies at one year. 	<p>Outcome Indicators:</p> <ul style="list-style-type: none"> ▪ fetal and infant deaths during perinatal period (28 weeks of gestation to 7 days or more after birth). ▪ all infant deaths (within one year). ▪ neonatal deaths (within the first 28 days of life). ▪ postneonatal deaths (between 28 days of life and one year). ▪ deaths due to Sudden Infant Death Syndrome (SIDS). ▪ maternal complications during hospitalized labor and delivery. ▪ low birth weight (LBW) infants. ▪ very low birth weight (VLBW) infants. ▪ preterm births. ▪ live births at 32 to 36 weeks of gestation. ▪ live births at less than 32 weeks of

INFANT MORTALITY			
Outputs (process outcomes)	Initial Outcomes	Intermediate outcomes	Long-term outcomes
		<ul style="list-style-type: none"> ▪ Pregnant women who receive early and adequate prenatal care. ▪ Mothers who achieve a recommended weight gain during their pregnancies. ▪ Newborn bloodspot screening, follow-up testing, and referral to services in the targeted minority populations. ▪ Reduce cesarean births among low-risk women (full-term, singleton, vertex presentation) ▪ 	<ul style="list-style-type: none"> gestation. ▪ Occurrence of neural tube defects (Spina Bifida, Anencephaly).

SUBSTANCE ABUSE PREVENTION			
Outputs (process outcomes)	Initial Outcomes	Intermediate outcomes	Long-term outcomes
<ul style="list-style-type: none"> ▪ Number of participants receiving educational intervention. ▪ Number of educational interventions offered. ▪ Number of hours of direct contact with program participants. 	<p>Outcome Indicators:</p> <ul style="list-style-type: none"> ▪ adolescents who: <ul style="list-style-type: none"> ○ disapprove of substance abuse. ○ perceive great risk associated with consuming five or more alcoholic drinks at a single occasion once or twice a week. ○ perceive great risk associated with smoking marijuana once a month. ○ perceive great risk associated with using 	<p>Outcome Indicators:</p> <ul style="list-style-type: none"> ▪ Percentage of program youth participants reporting no substance use at the beginning of the program who continue to report no substance use at the end of the program period ▪ Percentage of program youth participants reporting experimental use at the beginning of the program who now to report no substance use at the end of the program period 	<p>Outcome Indicators:</p> <ul style="list-style-type: none"> ▪ deaths caused by alcohol-related motor vehicle. ▪ injuries caused by alcohol-related motor vehicle crashes. ▪ deaths caused by drug-related motor vehicle crashes. ▪ injuries caused by drug-related motor vehicle crashes. ▪ cirrhosis deaths.

SUBSTANCE ABUSE PREVENTION			
Outputs (process outcomes)	Initial Outcomes	Intermediate outcomes	Long-term outcomes
	<ul style="list-style-type: none"> o cocaine once a month. o disapprove of smoking. ▪ participants showing improvement in knowledge/attitude/skills scores related to substance abuse (including alcohol and tobacco) prevention and cessation. 	<ul style="list-style-type: none"> ▪ Percentage of program youth participants who avoided alcohol and/or drug environments ▪ Percentage of youth participants who attend school at a 95% or better rate. ▪ Percentage of program youth participants who declined alcohol/and or drugs when offered to them ▪ Youth/Adolescents who report that they did not ride, during the previous 30 days, with a driver who had been drinking alcohol. ▪ Proportion of youth/adolescents never using substances ▪ Youth/Adolescents not using alcohol or any illicit drugs during the past 30 days. ▪ Youth/Adolescents reporting no use of marijuana during the past 30 days. ▪ Youth who report no engagement in binge drinking of alcoholic beverages during the past month. ▪ Youth/adolescents who report no Cigarette use in past month. 	<ul style="list-style-type: none"> ▪ drug-induced deaths. ▪ average age of first use of alcohol in adolescent's age 12-17 year old that remain alcohol free. ▪ average age of first use of marijuana in adolescents aged 12-17 year old that remain alcohol free. ▪ adolescents never using alcohol. ▪ adolescents never using illicit drugs. ▪ Number of admissions for substance abuse treatment for injected drug use.

VIOLENCE PREVENTION			
Outputs (process outcomes)	Initial Outcomes	Intermediate outcomes	Long-term outcomes
<ul style="list-style-type: none"> ▪ Number of participants receiving educational intervention. ▪ Number of educational 	<p>Outcome Indicators:</p> <ul style="list-style-type: none"> ▪ participants showing improvement in 	<ul style="list-style-type: none"> • Percentage of program participants who understand what violence is and their ability to identify violent acts • Percentage of youth participants who 	<p>Outcome Indicators - Reduced:</p> <ul style="list-style-type: none"> ▪ deaths caused by homicides. ▪ Fire-arms related injuries

VIOLENCE PREVENTION			
Outputs (process outcomes)	Initial Outcomes	Intermediate outcomes	Long-term outcomes
interventions offered. <ul style="list-style-type: none"> ▪ Number of hours of direct contact with program participants. 	knowledge/attitudes/skills scores related to violence prevention.	attend school at a 95% or better rate. <ul style="list-style-type: none"> • Percentage of program participants who report involvement in positive skill/social development activities • Percentage of program participants who report avoiding peers who have poor conduct. • Percentage of program participants who report avoidance of gang involvement or related activity. • Percentage of program participants who avoid potential incidents of physical altercations, bullying, harassing and/or electronic aggression 	and deaths. <ul style="list-style-type: none"> ▪ child maltreatment incidence. ▪ child maltreatment fatalities. ▪ the rate of physical assault by current or former intimate partners. ▪ rate of rape or attempted rape. ▪ rate of sexual assault other than rape. ▪ rate of physical assault. ▪ intentional injuries resulting from alcohol-related and illicit drug-related violence. ▪ Suicide. ▪ suicide attempts by adolescents. ▪ work-related assaults. ▪ Non-fatal intentional self-harm injuries

APPENDIX D.

II. GLOSSARY OF SELECTED EVALUATION TERMS

Activities are what a program does with its inputs—the services it provides—to fulfill its mission. Examples include conducting a screening program for breast cancer among Asian American community, educating the African American community members in a neighborhood about how to be physically active, and providing social support for sustaining physical activity in a group of Hispanic Americans. Program *activities* result in *outputs*.

Benchmarks are performance data that are used for comparative purposes. Baseline benchmark data is collected prior to the participants receiving the project activities. A program can use its own data as a baseline benchmark against which to compare future performance. It, also, can use data from another program as a benchmark. In the latter case, the other program often is chosen because it is exemplary and its data are used as a target to strive for, rather than as a baseline.

Community impact is a concept reflecting the collective effect of an intervention beyond the individual-level changes captured via specific outcome indicators. For example, a program seeking to prevent adolescent childbearing in specific neighborhoods, may, over time, be able to document declines in the teen pregnancy rate for the targeted geographic areas.

Inputs are resources a program uses to achieve program objectives. Examples include staff, volunteers, facilities, equipment, curricula, and money (funding and/or program budget). A program uses *inputs* to support *activities*.

Instruments are used to obtain information. The term, instrument, is often used interchangeably with the terms tool, survey, measurement, and questionnaire. It is also sometimes referred to as a scale.

Logic Model is a planning tool used to clarify and illustrate what your project intends to do and what it hopes to accomplish and impact. A logic model can be considered a map that you develop to clarify and communicate what your project intends to do and its presumed impact.

Outcomes are benefits for participants during or after their involvement with a program. Outcomes include changes in knowledge, skills, attitudes, values, behaviors, conditions or health status. Examples of outcomes include greater knowledge of nutritional needs, improved reading skills, more effective responses to conflict, getting a job and losing weight.

For a particular program, there can be various "levels" of outcomes, with initial outcomes leading to intermediate outcomes which then lead to longer-term outcomes. For example, a screening program for cervical cancer detection among African American women may show a 20% increase in Pap test for women over 18 years of age during a one year period. This program if sustained over a longer period of time and broadened in its reach would in turn lead to decreased death rates due to cervical cancer.

Outcome indicators are the specific items of information that track a program's success in achieving its outcomes. They describe observable, measurable characteristics or changes that

represent achievement of an outcome. For example, a program whose desired outcome is that participants pursue a healthy lifestyle could define "healthy lifestyle" as not smoking; maintaining a recommended weight, blood pressure, and cholesterol level; getting at least two hours of exercise each week; and wearing seat belts consistently. The number and percent of program participants who demonstrate these behaviors then is an *indicator* of how well the program is doing with respect to the outcome. Pre- and post-test surveys, attendance rosters, and numerous measurement tools are used as indicators to track a programs success.

Outcome targets are numerical objectives that specify a program's level of achievement on its outcomes. After a program has had experience with measuring outcomes, it can use its findings to set targets for the number and percent of participants expected to achieve desired outcomes in the next reporting period. It, also, can set targets for the amount of change it expects participants to experience. For example, an outcome target might state that 80% of participants will demonstrate knowledge gain from pre- to post-test survey results.

Outputs are products of a program's activities, such as the number of meals provided, classes taught, brochures distributed, or participants served. Another term for "outputs" is "units of service." A program's *outputs* should produce desired *outcomes* for the program's participants.

Participants refer to individuals who are partaking of program and/or services. Other terms may also be used to describe participants like consumers, attendees, or clients.

Reliability refers to the consistency of an instrument to yield the same results when the instrument is used to measure the same thing time and time again. An instrument is said to be reliable, therefore, when it consistently measures a concept or phenomenon with accuracy. Reliability asks the question: Do repeated applications under similar conditions yield consistent results? A reliable instrument is like a reliable watch – it is not subject to extraneous factors such as the race or weight or height or income of the person wearing it, the month of the year, day of the week, the temperature, etc.

Scale is a cluster of items intended to measure a single concept. For example, in a study designed to measure attitudes about diabetes, an instrument may contain 'knowledge of diabetes' scale, a 'fear of diabetes' scale, and 'action taken to prevent or manage diabetes' scale. An instrument may contain one or more scales depending on the variables that need to be measured in order to efficiently examine the concept being studied.

Validity is the correspondence between what a measuring device is supposed to measure and what it really measures. An instrument is said to be valid when it measures what it was designed/intended to measure.

APPENDIX E. ADDITIONAL RESOURCES

Cultural Competence

- Bennett, M. J. (1993). Towards ethnorelativism: A developmental model of intercultural sensitivity. In R. M. Paige (Ed.). *Education for the intercultural experience*. Yarmouth, ME: Intercultural Press.
- Campinha-Bacote, J. (1994). The process of cultural competence in health care: A culturally competent model of care. Cincinnati, OH: Transcultural C.A.RE. Associates.
- Deason, L. M. & Wallace, S. L. (2005). Does a shared culture between nursing home staff and residents influence quality of life? *International Journal of Diversity in Organisations, Communities, and Nations*, 4, 235-242.
- Harwood, A. (1981). Guidelines for culturally appropriate care. In A. Harwood (Ed.), *Ethnicity and medical care* (pp. 482-509). Cambridge, MA: Harvard University Press.
- Lenburg, C., Lipson, J.G., Demi, A.S., Blaney, D.R., Stern, P.N., Schultz, P.R., et al. (1995). *Promoting cultural competence in and through nursing education: A critical review and comprehensive plan for action*. Washington, DC: American Academy of Nursing.
- National Multicultural Institute. (1999). *Cultural competency in health care: A guide for trainers* (2nd ed.). Washington, DC: National Multicultural Institute.

Evaluation

- Allen, M. J., & Yen, W. M. (2001). *Introduction to measurement theory*. Long Grove, IL: Waveland Press.
- American Evaluation Association (AEA) www.eval.org
- Carmines, E. G., & Zeller, R.A. (1979). *Reliability and validity assessment*. Newbury Park, CA: Sage.
- Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. Newbury Park, CA: Sage.
- Crocker, L., & Algina, J. (1986). *Introduction to modern and classical test theory*. Fort Worth, TX: Harcourt Brace College Publishers.
- DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.

- Dignan, M. B. (1995). *Measurement and evaluation of health education* (3rd ed.). Springfield, IL: Charles C. Thomas.
- Fetterman, D. M., Kaftarian, S. J., & Wandersman, A. (1997). *Empowerment evaluation: Knowledge and tools for self-assessment and accountability*. Thousand Oaks, CA: Sage.
- Green, L. W., & Lewis, F. M. (1986). *Measurement and evaluation in health education and health promotion*. Mountain View, CA: Mayfield.
- Hand, D. J. (2005). *Measurement theory and practice*. London: Arnold Publishers.
- Juvenile Justice Evaluation Center (2001). *Hiring and Working with an Evaluator*. Available at <http://www.jrsa.org/jjec/about/publications/evaluator.pdf>
- Kline, T. J. B. (2005). *Psychological testing: A practical approach to design and evaluation*. Thousand Oaks, CA: Sage.
- Leech, N. L., Barrett, K. C., & Morgan, G. A. (2005). *SPSS for intermediate statistics: Use and interpretation*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Lorig, K., Stewart, A., Ritter, P., Gonzalez, V., Laurent, D., & Lynch, J. (1996). *Outcome measures for health education and other health education interventions*. Newbury Park, CA: Sage.
- Miller, D.C. (1991). *Handbook of research design and social measurement* (5th ed.). Newbury Park, CA: Sage.
- Morgan, G. A., Leech, N. L., & Barrett, K. C. (2005). *SPSS for introductory and intermediate statistics* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, CA: Sage.
- Nunnally, J. C., & Bernstein, I. (1994). *Psychometric theory*. New York: McGraw Hill.
- Ohio Program Evaluators' Group (OPEG) www.opeg.org.
- Patton, M. Q. (2001). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Pedhazur, E., & Schmelkin, L. (1991). *Measurement, design, and analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. Thousand Oaks, CA: Sage.

- Polit, D. F., & Beck, C. T. (2003). *Nursing research. Principles and methods* (7th. ed.). Philadelphia, PA: Lippincott Williams & Wilkins.
- Rogers, H. J., Swaminathan, H., & Hambleton, R. K. (1991). *Fundamentals of item response theory*. Thousand Oaks, CA: Sage.
- Sarvella, P.D., & McDermott, R.J. (1993). *Health education and measurement: A practitioner's perspective*. Madison, WI: Brown & Benchmark.
- Shultz, K. S., & Whitney, D. J. (2004). *Measurement theory in action: Case studies and exercises*. Thousand Oaks, CA: Sage.
- Streiner, D. L., & Norman, G. R. (2003). *Health measurement scales: A practical guide to their development and use*. Oxford University Press.
- Tabachnick, B. G., & Fidell, L. S. (2000). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn & Bacon.
- Viswanathan, M. (2005). *Measurement error and research design*. Thousand Oaks, CA: Sage.
- Waltz, C. F., Strickland, O. L., & Lenz, E. R. (2004). *Measurement in nursing and health research*. New York: Springer.
- Wholey, J.S., Hatry, H.P., and Newcomer, K.E. (1994). *Handbook of practical program evaluation*. San Francisco, CA: Jossey-Bass.

Health Disparities

- Association of State and Territorial Health Officials, National Association of County and City Health Officials, and HRSA (undated) *Health Departments Take Action. A compendium of state and local models addressing racial and ethnic disparities in health*. Retrieved January 31, 2006 from www.astho.org/pubs/nabookfull.pdf
- Berkman N.D., DeWalt D.A., Pignone M.P., Sheridan S.L., Lohr K.N., Lux L., et al. (2004). *Literacy and health outcomes*. Evidence Report/Technology Assessment No. 87 (Prepared by RTI International–University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016). AHRQ Publication No. 04-E007-2. Rockville, MD: Agency for Healthcare Research and Quality.
- Institute of Medicine. (2002). *Unequal treatment: Confronting racial and ethnic disparities in health care*. Washington, DC: National Academy Press.
- Randall, V. (2005) *Race, Health Care, and the Law*. Available at <http://academic.udayton.edu/health/index.htm>

United States Department of Health and Human Services. (2000). *Healthy People 2010*. Washington, DC: U.S. Government Printing Office.

Health Literacy

Agency for Healthcare Research and Quality. (2004). *Literacy and health outcomes*. Report Number 04-E007-1.

Bresolin, L. (1999). Health literacy. *The Journal of the American Medical Association*, 281, 552-557.

Davis, T. C., Frederickson, D. D., Arnold, C., Murphy, P. W., Herbst, M., & Bocchini, J. A. (1998). A polio immunization pamphlet with increased appeal and simplified language does not improve comprehension to an acceptable level. *Patient Education Counsel*, 33, 5-37.

Institute of Medicine. (2004). *Health Literacy: A prescription to end confusion*. Washington, DC: The National Academy Press.

Michielutte, R., Bahnson, J., Dignan, M.B., & Schroeder, E. M. (1992). The use of illustrations and narrative text style to improve readability of a health education brochure. *Journal of Cancer Education*, 7, 251-260.

Murphy, P. W., Chesson, A.L., Walker, L., Arnold, C. L., & Chesson, L. M. (2000). Comparing the effectiveness of video and written material for improving knowledge among sleep disorders clinic patients with limited literacy skills. *Southern Medical Journal*, 93, 297-304.

National Library of Medicine. (2000). *Current bibliographies in medicine: Health literacy*. Bethesda, MD: National Institutes of Health.

Williams, M.V., Baker, D.W., Honig, E.G., Lee, T.M., & Nowlan, A. (1998). Inadequate literacy is a barrier to asthma knowledge and self-care. *Chest*, 114, 1008-1015.

Williams, M.V., Baker, D.W., Parker, R.M., & Nurss, J.R. (1998). Relationship of functional health literacy to patients' knowledge of their chronic disease. A study of patients with hypertension and diabetes. *Archives of Internal Medicine*, 158, 166-172.

Multicultural Evaluation

California Endowment. (2005). *Commissioning Multicultural Health Evaluation: A Foundation Resource Guide*. Retrieved February 1, 2006 from www.calendow.org/reference/publications/pdf/evaluations/TCE0510-2004_Commissioning_.pdf

California Endowment. (2003). *Overview of Multicultural and Culturally Competent Program Evaluation: Issues, Challenges, & Opportunities*. Available at

www.calendow.org/reference/publications/pdf/evaluations/TCE0509-2004_Overview_of_Mu.pdf

California Endowment. (2003). *Toward Culturally Competent Evaluation in Health and Mental Health*. Available at

www.calendow.org/reference/publications/pdf/evaluations/TCE0511-2004_Toward_Cultura.pdf

California Endowment. (2003). *Foundation Leaders' Perspectives on Culturally Competent Evaluation*. Available at

www.calendow.org/reference/publications/pdf/evaluations/TCE0508-2004_Foundation_Lea.pdf

California Endowment. (2003). *Voices from the Field: Health and Evaluation Leaders on Multicultural Evaluation*. Available at

www.calendow.org/reference/publications/pdf/evaluations/TCE0515-2005_Voice_From_The.pdf

Georgetown University Center for Child and Human Development. (2003). *A Guide to Choosing and Adapting Culturally and Linguistically Competent Health Promotion Materials*.

Retrieved February 1, 2006 from Available at:

www.gucchd.georgetown.edu/nccc/documents/Materials_Guide.pdf.

. U.S. Department of Health and Human Services, Office of Civil Rights. (2001) *Policy Guidance Title VI Prohibition Against National Origin Discrimination As It Affects Persons With Limited English Proficiency*. Available at: www.hhs.gov/ocr/lep/guide.html

Reliability and Validity

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Dilorio, C. (2005). *Measurement in health behavior and health education: methods for research and education*. San Francisco: Jossey-Bass.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: issues and applications*. Thousand Oaks, CA: Sage.

Schnall, J. G., & Rich, J. (1994). Finding measurement tools. Retrieved October 21, 2005 from <http://healthlinks.washington.edu/howto/measure.html> .

Streiner, D. L., & Norman, G. R. (2003). *Health measurement scales: A practical guide to their development and use*. New York: Oxford University Press.

University Library at University of Illinois at Chicago. (2004). Health instruments. Available at <http://www.uic.edu/depts/lib/lhsu/resources/hlthinst.shtml> .

Selection of an Evaluator

Juvenile Justice Evaluation Center. (2001). *Hiring and working with an evaluator*. Available at <http://www.jrsa.org/jjec/about/publications/evaluator.pdf>

Stufflebeam, D.L. (1999) Evaluation contracts checklist. Available at <http://www.wmich.edu/evalctr/checklists/contracts.htm>

2012 Revisions to this document were made by current REEP panelists:
Joann Wright Mawasha, Psy.D., Wright State University, Panel Chair
Rob Fischer, Ph.D., Case Western Reserve University
Peter Leahy, Ph.D., University of Akron
Fatima Perkins, MSLIS, M.N.O., Consultant
Manoj Sharma, Ph.D., University of Cincinnati

2016 Revisions to this document were made by current REEP panelists:
Joann Wright Mawasha, Psy.D., Wright State University, Panel Chair
Fatima Perkins, MSLIS, M.N.O., Consultant
Manoj Sharma, Ph.D., University of Cincinnati
Raghav Suri, Psy.D.