

**CHAPTER TWO:
PRINCIPLES OF PERFORMANCE
MEASUREMENT**

<u>TOPIC</u>	<u>Page</u>
Principles of Performance Measurement	2-3
Principle 1: Know What Performance Measurement Is	2-3
Principle 2: Useful, Feasible, Accurate, and Respectful.....	2-6
Principle 3: Begin with a Logic Model for Your Program	2-9
Principle 4: Know Your Capacity For Assessment	2-13
Principle 5: Know Your Design: Who, What, When, and How	2-19
Principle 6: Use What You Learn.....	2-23
Some Final Words.....	2-27

Chapter Two:

Principles of Performance Measurement

Performance measurement is an integral component in accountability. Despite its complexity, six basic principles guide performance measurement and assessment of outcomes.

Principle One: Know What Performance Measurement Is

Performance measurement determines the success of a specific program or initiative by comparing plans to actual activities and outputs, and outcomes. Performance measurement is characterized by its inclusion in ongoing operations as a part of every day management. Measures are used that both inform service delivery and demonstrate results.

Performance measurement answers critical questions, such as:

What are the planned resources, activities and outcomes?

What resources are actually used?

What is actually done by a program or initiative?

Who actually participates?

Are the anticipated outcomes or results reached?

Do some participants benefit more than others?

What do these efforts and results mean for future programs?

Performance measurement analyzes the success of program efforts by comparing data to what actually happened to what was planned or intended. Performance measurement requires a clear vision of what is planned and systematically gathered information about what actually occurs.

This information usually comes from staff, program records and observations, and includes the resources expended, the numbers and types of activities conducted, and the numbers of participants, and the outcomes for participants.

Participant characteristics may also be important in analyzing results. These include gender, family structure, income level, risk status, age or other characteristics of interest to the program.

PERFORMANCE MEASUREMENT ESSENTIALS

- *A clear vision of intended activities and results*
- *Systematically gathered information about actual participants, activities and outcomes*
- *Careful analysis of efforts and outcomes to inform future decision-making*

Performance measurement focuses on program implementation and tracks intended inputs, activities, and outputs on a regular basis.

Does the program or initiative do what we set out to do?

- *When we compare our planned inputs and outputs with what our actual activities did, did we do what we set out to do?*
- *Did the people we intended to participate, do so? Who did participate?*
- *Did participants like the program activities? Were they treated with respect? Were their service needs met?*

In addition, performance measurement also answers questions about outcomes or results. Sometimes these questions ask about ways in which different participants may have benefited from the program.

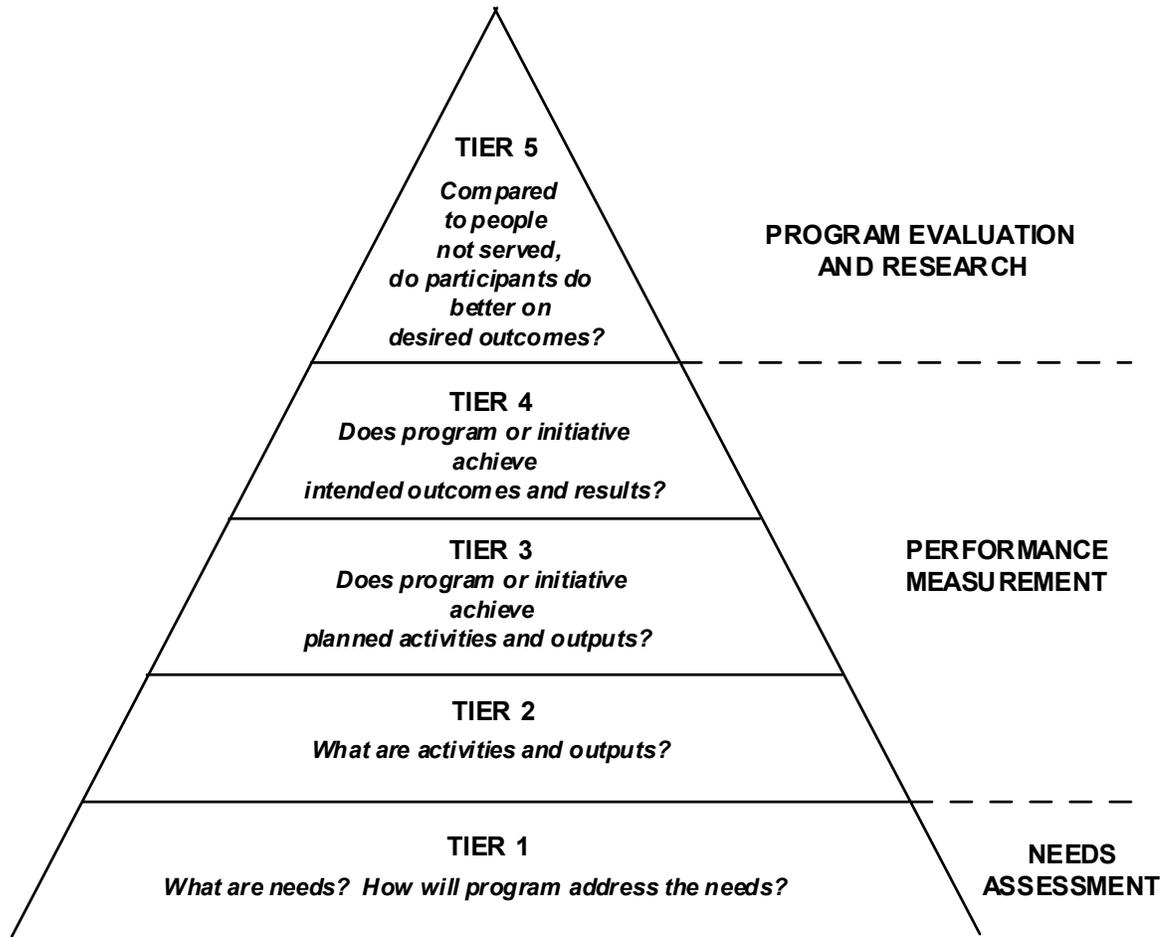
Evidence must be systematically gathered to answer questions about outcomes or results. This evidence may be collected throughout the period of service delivery or only at the end of the program.

Does the program achieve its intended outcomes or results?

- *Do clients report positive results from our program?*
- *Do clients reach the outcomes set for the program?*
- *Do clients improve? (pre/post or retro-pre/post)*
- *Do some clients benefit more (or less) than others?*
- *If so, what are the characteristics of people who benefit most? (age, gender, marital status, initial level of needs, or other characteristics)?*
- *Do the activities (attendance, services received) for participants who benefited differ from the activities for those who did not benefit as much?*

All *program evaluations* build on the information gathered in performance measurement. In one widely used five-tier evaluation model for family programs, performance measurement comprises *three* of the five tiers (Figure 2-1). But only the fifth and highest tier addresses the fundamental *evaluation* question of causality.

Figure 2-1: Tiers of Evaluation and Performance Measurement



SOURCE: Weiss, H.B. & Jacobs, F. Evaluating Family Programs.

Performance measurement enhances and complements in-depth evaluation (Hatry, 1997). But to establish causality, evaluation requires all the information that is gathered in the performance measurement tiers PLUS information on *outcomes for a similar group* of people who *did not participate*. This is usually called a comparison or control group.

In evaluation designs that seek to establish causality, people in the control group are identical to those who are served. Further, everyone (participants and people in the control group) are randomly assigned to

Principles

ensure comparability between those served and those not served by a particular intervention or “treatment.”

Program evaluations are more complicated and expensive than performance measurement. Evaluations require not only tracking who participates, what activities take place, and what outcomes are achieved, but also comparing these results in a *scientific manner* to a *similar group* of people who *did not participate* in the program.

Questions addressed by program evaluations may include:

- Are program results due only or mostly to the program itself?
- Compared to people who were not served, do program participants do better on the desired outcomes?

Program evaluation *is essential* to advance knowledge and practice. Not all programs need to use precious resources to conduct evaluation and research. Instead, programs can effectively build on what is learned from carefully controlled evaluations and research.

This is the idea behind *Building Results I* that summarizes program outcomes and activities, demonstrated by research and evaluation to be related to Oregon’s long-term goals and Benchmarks. Using *Building Results I*, programs can select activities and outcomes which will lead to Oregon’s wellness goals and Benchmarks.

Building Results I
Uses
Evaluation and Research

to identify outcomes and activities for individual programs and initiatives

that will help Oregon reach its Benchmarks and wellness goals.

Almost all programs or initiatives can build on best practices and can use performance measurement strategies to track their inputs, activities and outputs, and results. But every program or initiative does not have the resources to do full program evaluations.

Principle Two: Useful, Feasible, Accurate, and Respectful

Performance measurement describes what a program does, whom it serves, and what it accomplishes. When a performance measurement system is established, many decisions have to be made such as what to record, who to get information from, and when and how to gather information.

What guidelines can you use to help make these decisions? What are the characteristics of a good performance measurement system?

When asked these questions, managers and evaluators often respond that a good performance measurement system provides useful and accurate information that is collected in a feasible and respectful manner.¹

Useful information. Good performance measurement provides information that is timely and relevant to service planning and delivery. Useful reports should summarize findings and be accessible for decision-making. Most importantly, useful information is able to **show the impact** of your program or initiative during its time of operation.

Keys to A Good Performance Measurement System

- Useful,
- Accurate,
- Feasible, and
- Respectful

Accurate information. Good performance measurement provides information that is believable and correct. Accurate information builds on valid and impartial standards, reliable procedures, and reasonable interpretations and conclusions.

GOOD OUTCOME MEASURES

- *Can demonstrate results during the time of service.*
- *Are relevant, accurate, and believable.*
- *Help identify strengths and short-comings in the program.*
- *Measure outcomes that are valued by stakeholders, including decision-makers.*

A limited amount of accurate information is better than a lot of inaccurate and incomplete information. Effective performance measurement strives to gather information that is accurate.

Feasibility. Good performance measurement uses resources realistically and wisely. This means that information is gathered in a manner that is manageable over time. Careful planning can assure that measurement strategies are practical and do not add unnecessary work or record-keeping.

For performance measurement to be feasible, staff who are responsible for data collection, maintenance, and analysis must have the time, skills, and resources to do this important task. In choosing measures, it is critical that the skills and resources required be available. Otherwise, the information that is collected is likely to be incomplete and inaccurate.

¹ These same four standards guide evaluation research as well. In 1981 a group of researchers, evaluators, and educators, joined to form the Joint committee on Standards in Educational Evaluation. The four standards were adapted: Utility, Accuracy, Feasibility, and Propriety. In 1994, these standards were reaffirmed and are now adopted by 15 professional organizations including the American Evaluation Association, the American Psychological Association, the American Educational Research Association, and others.

In judging what is feasible, it is critical to balance the value of the information that will be gathered to the costs (staff time, money, and other resources) of collecting that information.

As a wise person once said:

“Inaccurate information that is easy to get is useless. Great information that is impossible to get is also useless.”

One cost to consider is the opportunity costs – when a staff person is collecting information it is likely that some other task is not being done. When the information that is collected directly *benefits* participants and program operations, staff are more willing to invest time and resources in gathering that information.

Thus, most service programs try to build their outcomes measurement on program records, observations, and other assessments that support service delivery. In short, no matter how feasible some measure may be to use, only useful and accurate information is worth collecting.

Respectful. Good performance measurement provides information that represents different interests and provides a complete and fair assessment. In addition, respectful performance measurement:

- Addresses worthwhile (not just measurable) outcomes, and
- Protects the rights and welfare of participants and involved staff, including respecting confidentiality, dignity, time, and other non-service needs.

{See Chapter 4, pages 4-20 to 4-22 for more on confidentiality and informed consent.}

Finally, respectful processes acknowledge the challenges faced by communities and programs as they expand performance measurement to include assessment of outcomes. Most communities and programs have little or no experience with outcome assessment. They need time and support to develop their capacity for full performance assessment. Leadership, patience, and a willingness to experiment, are essential to build a performance measurement system.

“Respectful leadership and a willingness to experiment are essential to build a performance measurement system.”

Principle Three: Begin with a Logic Model for Your Program

Performance measurement compares what was intended to what occurred. Thus, the first critical step in performance measurement is the creation of a **logic model** that clarifies the exact inputs, activities and outputs, and outcomes planned for a program or initiative.

A logic model shows the “chain” of activities and outputs that are intended to lead to the program outcomes. Usually, logic models are created in several steps:

Step one. Define the long-term goal and/or benchmark that you want to reach – write this on the far right side of a big piece of paper!

LOGIC MODELS
show how activities will lead to outcomes and, eventually, to long-term goals.

Step two. On the left side of the paper, write the basic components of the program or initiative that will be offered. Be sure these components reflect essential best practices and proven approaches that are known to lead to the outcomes you want to reach.

Step three. Answer the question “what is the first (then second, then third or more) outcome that should occur when this program is implemented?” This is sometimes called the “**so that**” chain. Research that identifies the knowledge, skills, attitudes and behaviors that lead to long term goals and benchmarks can help you define the outcomes in a **so that** chain.

For example, to contribute to reaching the long-term goals of positive youth development and educational success, a school-based conflict resolution program offers training to all students and staff. This program builds on what has been proven to be effective, such as on-going classroom instruction, practice for skill building, and application in *all* conflicts by *both* students and staff (Johnson & Johnson, 1995; 1996). This conflict resolution program is implemented **so that**:

- students and staff all know the steps in conflict resolution, **so that**
- they can demonstrate these steps in role-play situations, **so that**
- they can utilize these steps in conflicts that arise in school, **so that**
- student conflicts that require teacher or other adult intervention decrease, **so that**
- referrals for student to student conflicts decrease, **so that**
- school becomes a more positive environment for all youth, **so that**
- educational success is enhanced.

The knowledge, skills, attitudes, and/or behaviors in the *so that chain* are the outcomes or results of the program. A manager will have to decide how far up the chain to measure outcomes. The further up the chain, the more other factors influence an outcome.

“So That” Chains

Use research to align program activities and outcomes with long-term goals and Benchmarks.

Define outcomes that are under the direct influence of the program or initiative.

For example, in the conflict resolution training program, the final outcomes are that “school is a more positive place” and “educational success is enhanced.” These outcomes are clearly influenced by many factors including adequate funding, skilled educators, effective curriculums, and family support.

Many different programs or initiatives will be needed to achieve the important long-term goal of educational success for all children and youth. These programs or initiatives must be coordinated to effectively address different aspects of the long-term goal.

In addition to effective school-based conflict resolution, other contributors to positive school climate and educational success include:

- adequate facilities, equipment, and resources,
- effective instruction and support for children at all skill levels, and
- parent and community support and participation.

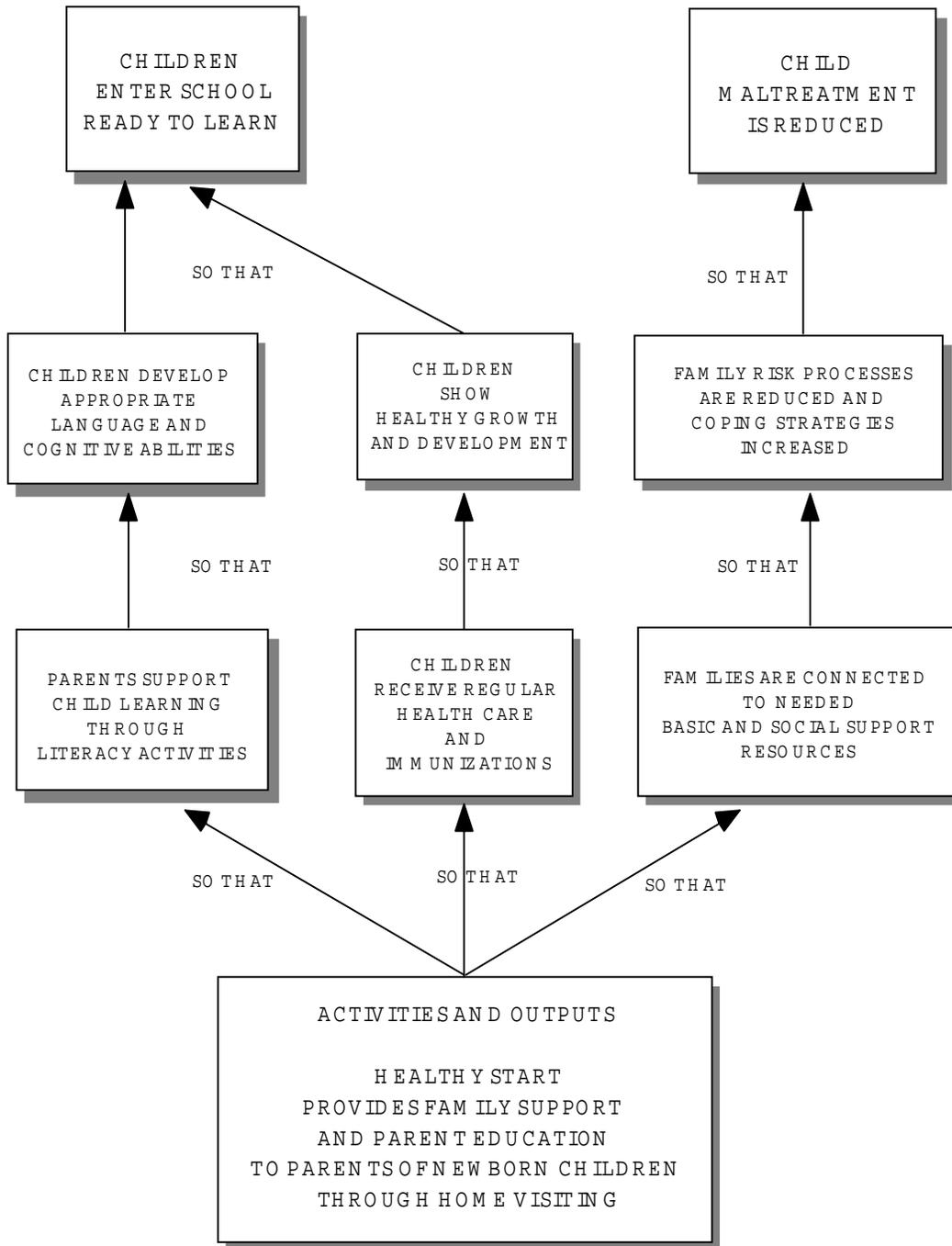
Thus, the goals of positive school climate and educational success should *not* be measured as evidence of its success of a single program. Given this, the conflict resolution program should *assess the outcomes it most directly affects*, such as:

- increased knowledge and skills in conflict resolution,
- reduced referrals for student to student conflicts, and
- reduced need for adult intervention in student to student conflicts.

“Many programs and initiatives are needed to reach an important long-term goal such as educational success for all children.”

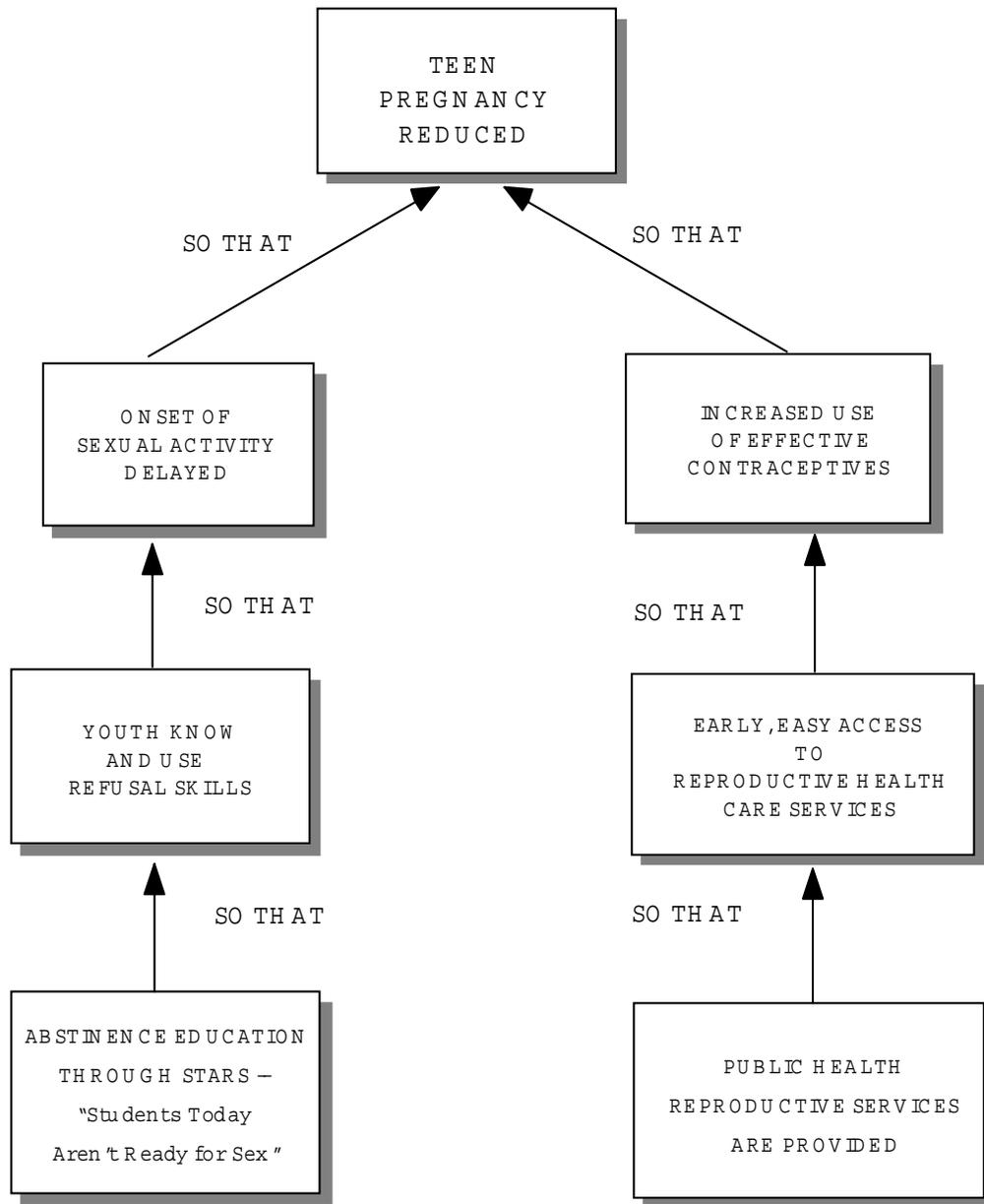
More examples and information logic models are found in *Bridging the Gap: Building Results II* (OCCF, 1997) and the United Way of America publication (1996) *Measuring Program Outcomes: A Practical Approach*. In addition, sample logic models are presented on the next two pages.

Figure 2-2: Logic model for Oregon Healthy Start



Sometimes a comprehensive effort, such as Oregon Healthy Start, works with several other agencies and organizations to address more than one long-term goal.

Figure 2-3: Logic Model for Teen Pregnancy Prevention Efforts



Sometimes a single long-term goal is addressed through several efforts. The Oregon Governor's Teen Pregnancy Prevention Agenda identifies five initiatives to reduce teen pregnancy. Two of these, STARS and reproductive health care access and utilization are shown above.

Principle 4: Know Your Capacity For Assessment

It is never a good idea to try to do more than you really can. In considering capacity, you have to make many decisions. Trade-offs will be made between what would produce the best, most accurate information and what is actually possible and respectful to do. In short, accuracy, feasibility and respect have to be balanced.

For example, the *most accurate* information on parent-child interactions is likely to come from trained observers videotaping parent-child interactions for several days in different situations. Subsequently these video tapes would be carefully coded by several highly trained raters.

This method, however, is likely to be too invasive for most families. It is also very expensive in terms of time, money, and technical resources. Thus this method, while very accurate, does not meet the respectful and feasible standards of performance measurement (see Principle 2).

BALANCE TRADE-OFFS

Ask yourself:

What will produce the most accurate information?

What is actually feasible to do?

What is respectful of participants and staff?

A trade-off would be for parent-child interaction information to be gathered in some other valid, reliable way, such as:

- Trained staff members rate interactions using the "Interactions between Caregivers (Parents)/Children" and the "Developmental Stimulation of Children" scales of the Family Assessment Form (FAF) (see Tables 5-1 and 5-10)
- Trained staff members rate interactions using the Parent-Child Interaction Scale of the Oregon Healthy Start Evaluation (see Table 5-10)
- Parent and/or children complete the Parent-Child Relationship Survey, the Parental Acceptance-Rejection Questionnaire, or other self-reports on relationships (see Table 5-10)

VALUE

A little accurate and complete information is better than a lot of inaccurate or incomplete information.

These options will provide accurate information that is sufficient to make program decisions while respecting the time of participants and recognizing resource limitations.

In all situations, a smaller amount of accurate and complete information is better than a lot of inaccurate or incomplete information. While this seems obvious, when programs first attempt to assess outcomes, they often over-reach and try to do too much.

Chapter 3 provides an overview of assessment approaches that range from simple (usually those based on records and information that is already collected in order to provide services) to more complicated methods (interviews and scales that demand staff training to administer.) Every assessment approach demands resources and has unique advantages and disadvantages. These are reviewed in greater detail in Chapter 3.

Self-assessment can help a program to define its capacity for performance measurement. Self-assessment ensures that chosen performance measurement strategies really fit that program's resources and participants.

A simple self-assessment tool is presented on the next page (Figure 2-5). This self-assessment is focused on clarifying a program's capacity for performance measurement. In this self-assessment, a higher score is not better...it is just different.

Completing this self-assessment can help to clarify capacity. This self-assessment includes questions about program history, intensity, participants, information needs, and resources. Based on these, the self-assessment rates a program's overall capacity to do performance measurement, especially outcome assessment.

Once a program has realistically defined its capacity for assessment, it can better select appropriate measures for its outcomes and choose a realistic design for collecting information with these measures. Chapter 3 provides more information about selecting measures; Principle 4 addresses design in performance measurement.

REMEMBER

***be clear about your
capacity***

***develop a measurement
system that fits....***

KEEP IN MIND

*"just because a ten year old
child can do **more** than a
three year old child,
that doesn't mean the ten year
old is better"*

Figure 2-4

Self-assessment: Performance measurement capacity

Circle the number that **BEST** describes your program or initiative:

How established is your program or initiative?

How long has your program or initiative been operating? What is its size, stability of staff and leadership, community support, and funding level?

1. Really just getting started – operating less than 3 years and many elements are still being developed.
2. Established for 3-5 years and most elements are working pretty well.
3. Established over 5 years and working smoothly; funding, activities and leadership are in place and pretty stable.

How intense is your program?

How frequent and how intense are contacts with participants?

1. Most contacts pretty brief; one-to-one contact is rare; most contact is in group settings; total contact time is less than 3 or 4 hours.
2. Most participants see us one-on-one and in group settings for at least 5 hours.
3. Extensive contacts with participants. Usually several contacts in small groups or extensive one-on-one contact.

How much and how complicated is the information that you need?

1. Only basic information on our program (who we serve, what we do) and one or two basic, immediate outcomes.
2. Basic program information plus information about several outcomes or about outcomes that occur over longer periods of time.
3. Basic information plus information on more complicated or extensive outcomes that are more difficult to achieve.

What resources do you have for performance assessment?

What money, staff skills and time, equipment, and technical assistance are available for data collection, record-keeping, and analysis?

1. Limited resources of all kinds
2. Adequate resources if we make it a priority
3. Good, reliable resources and high priority

SCORING: The most important part of scoring this self-assessment is THINKING!

What number did you circle most often?

If you circled a mix of 1, 2 and/or 3, then use your professional best judgement to assign a score. For example, if a program has never done outcome measurement, score its capacity lower, than if it has prior experience.

Adapted from: Goddard, W. et al. (1994). *The Alabama Children's Trust Fund Evaluation Manual*. Auburn University, Auburn, GA.

Principles

Higher scores on this self-assessment (Figure 2-4) generally indicate the capacity to use more complicated measures and approaches to track inputs, activities and outputs, and outcomes.

Level one capacity is limited to the most simple of data collection methods. These programs should probably rely on careful program records and simple assessments conducted at the end to the program to identify outcomes.

For example, at the end of the program participants could complete a simple interview or questionnaire describing how much the program helped them (if at all). Such post-program assessments are stronger if the participants are asked to rate themselves relative to the specific outcomes desired for the program. Participants could also be asked to rate how they were before the program and after. (This is called a retrospective pre-post test and is described more in the next section, under Principle 5.)

Level two capacity for performance measurement means that programs are able to collect *all the level one information and more*. For example, at level two, a program might add trained staff observations to outcome assessments at the end of the program or compare pre-post observations or surveys. Simple goal attainment scaling is also a level two strategy.

Level three capacity exceeds levels one and two. At level three, programs are ready to assess multiple outcomes, use multiple assessments, and/or use more complicated approaches such as observation scales, goal attainment scales, or interviews. Some of these may be collected in a

“pre-post” design which requires more complicate record-keeping and analysis.

YOUR PARTICIPANTS

1. *Most would need a lot of help with even a basic questionnaire*
2. *Most could complete a simple questionnaire*
3. *Most could complete a comprehensive questionnaire*

At every level of capacity, consider your participants in selecting your measurement approach. Your organization’s capacity may be at a level 3 (or higher!) but you still need to consider your participants’ capacity as you select your measures.

If your measures don’t fit your participants, then those measures won’t give you accurate information! Page 3-17 in the next chapter discusses choosing methods for your audience.

An example of how the Linn, Benton, Lincoln Service Integration Project approached outcome assessment is shown on the next two pages. Notice that they collected information from records as well as participants and others.

An Example of Performance Assessment

Linn- Benton- Lincoln Counties Service Integration Project

Project Overview

Through a network of interagency Youth Service Teams in the three counties, this Service Integration project offers in-home case management to higher risk families and youth. Interventions also include information and referral for basic resources, health and mental health care, ATOD services and other supports.

Five Family Service Coordinators (FSC) visit with families in their homes, attend school and agency meetings with these families, and assist families to set goals and develop plans for reaching these goals. Each FSC serves approximately 25 families intensively and 30 families with shorter term support and referral.

Community partners include county health and juvenile departments, Adult and Family Services, Services for Children and Families, law enforcement, local school districts, Community Services Consortium, local non-profit providers, and local Commissions on Children and Families.

The long-term vision for this project is to

- increase families' access to health and social services and
- enhance the resiliency of children and youth by decreasing risk factors and increasing protective factors in their families, schools, and communities.

Selected Specific Outcomes: (only two of five specific project outcomes are shown here)

- 1. 80% of families will be successfully linked to needed health and social services.**
- 2. 80% of students receiving long-term support will show improvement in social and academic performance.**

Outcome Assessment:

A *level one* assessment strategy was used to track these outcomes in the middle school component of this project. Level one assessment utilizes program records and simple assessments conducted at the end of the program to identify outcomes (see pages 2-16 and 2-17).

Linkage of families to needed health and social services (outcome 1) was assessed with program records and a “customer satisfaction survey.”

- Records indicated that in six months, the program outputs included over 1000 immunizations, 54 school-based health clinics, 576 nurse assessments, 175 counseling groups and 422 counseling sessions, and 203 referrals.
- Surveys indicated that 100% of parents and 78% of youth reported increased access to, and receipt of, needed health and/or social services.

Improvements in students’ social and academic performance (outcome 2) were assessed with a "Perception of Student Progress Scale." Using this scale at the end of three months of service, students and staff rated each student’s progress in six areas. Responses ranged from 1 (worse) to 4 (much improved). A “don’t know” response was included. During a six-month period, sixty-four youth completed self-assessments. Staff completed assessments on 56 youth.

The percentage of youth reported to have made improvements in each of the six areas ranged from 20% to 80% (see below). Students least frequently reported improvements in family relationships (45%). Students' most frequently reported improvements were in the areas of “controlling my behavior” and achievement at school. Staff reported improvements least frequently in family adjustment (20%) and the most frequently in social skills (80%) and emotional adjustment (75%).

Youth self-reporting improvements	
67%	controlling behavior
62%	handling emotions
45%	getting along at home
51%	getting along with peers
56%	getting along at school
66%	school grades

Staff reporting youth making improvements	
71%	behavioral control
75%	emotional adjustment
20%	family adjustment
80%	social relationship skills
70%	school adjustment
59%	school achievement

Principle Five: Know Your Design-Who, What, When, and How

Based on a program’s capacity for performance measurement, an appropriate measures and design for collecting information can be selected. The *design* is the master plan for conducting research, evaluation, or performance measurement. At a minimum, the design specifies:

- When, how, and what information or data will be collected;
- From whom information or data will be collected;
- How information or data will be analyzed; and
- How participants will be treated and their rights and welfare protected.

Some designs are very complicated. These designs are used in research and evaluation in order to understand causality by answering the question: “Does this service or treatment cause a particular outcome?” The most rigorous design for research or evaluation is a “pure experiment.”

Unlike such a rigorous experimental design, performance measurement does not seek to prove causality. Rather, performance measurement assesses how well a program implemented its plans and achieved its intended outcomes.

Performance measurement must fit in the “real world” of service delivery. Hence, performance measurement strategies rely on simpler designs such as pre-post tests, retrospective pre-post tests, post-test only. Records, staff observations and ratings, and self-reports are commonly used in performance measurement.

Participants are rarely randomly assigned to treatment or service conditions. Rather, the priority is given to serving those in need. In fact, program staff often feel it is unethical to deprive people of a program simply by chance.

***THE MOST RIGOROUS
EVALUATION DESIGNS***

- *Randomly sample and assign similar people to various treatments/ services,*
- *Define and consistently apply all treatment conditions,*
- *Utilize multiple, sophisticated measures and collect information at multiple points in time, and*
- *Utilize complicated statistical analysis.*

Simple designs are typically used in performance measurement. The most common performance measurement designs are:

SIMPLE DESIGNS

are typical in performance measurement.

- Post-Program Only
- Retrospective Pre-Post-Program
- Pre-Program, Post-Program
- Pre-Program, Post-Program, Long-term Post-Program

These designs are described further in Table 2-6.

Community development and systems change initiatives can be assessed using these same basic designs. Information about community conditions can be described before and after an initiative, using a combination of records, interviews, and observations.

Suppose a community initiative was designed to reduce conditions that promote under age drinking. Before (pre) and after (post) the initiative, information on community conditions is collected. This information might be collected through interviews with key informants (agency and school administrators, advocates, youth) and review of community data (number of community-youth events that are alcohol free; number of outlets cited for providing alcohol to minors; arrest and conviction rates for alcohol-related offenses among youth and adults).

Similarly, a “systems change effort” aimed at developing a “family friendly” orientation in community agencies could assess, *before and after* the initiative, the use of:

Community initiatives can also use similar basic designs to assess outcomes.

- Shared intake procedures and records;
- Procedures and services that involve families in decision-making;
- Co-located services and active interagency referral; and
- Shared outcome assessment procedures.

Community initiatives are almost always assessed on a “case study basis.” Rather than compare one community to another, a community is compared to itself using a pre-post design. In these assessments, the fundamental question becomes: “Is this community improving?” (Plantz, Greenway, & Hendricks, 1997).

Figure 2-5: Measurement Designs and Resources

Name of Design:	Description	Resources
Post-Program Only	Describe participants' behavior, attitudes, knowledge, skills, and/or circumstances <i>following</i> a program.	Low
Retrospective Pre- and Post-Program OR Post-Then-Pre Test	<p><i>Following</i> a program, participants' describe their behavior, attitudes, skills, knowledge and/or circumstances as they were <i>prior to</i> and <i>after</i> a program.</p> <p>Retrospective pre-tests measure change more accurately when participants' limited information before a program reduces their ability to correctly assess their behavior (Rockwell & Kohn, 1989; Linn & Slinde, 1977).</p> <p>Example: On a traditional pretest, participants indicated that they ate "one food rich in vitamin C daily." During the program they may learn that what they thought had vitamin C, does not.</p> <p>Example: On a pre-test, youth rated themselves as "able to work hard." After eight weeks of making trail in the rain, "hard work" had a different meaning for these youth. Their post-program scores actually dropped. In contrast, their adult supervisors reported significant improvements in the youths' work attitudes and behaviors, including the ability to "work hard." (See Table 8-6).</p> <p>Page 4-12 presents an example of a retrospective pre-post test.</p>	Low to Moderate
Pre- and Post-Program	Describes participants' behavior, attitudes, skills, knowledge and/or circumstance <i>prior to</i> and <i>after</i> a program . Requires managing information collected at two points in time.	Moderate
Pre-, Post-, and Long-Term Post Program	Same as pre- and post-program measure approach, with additional scores <i>obtained again at a later point in time</i> (e.g., six months, one year, two years).	High

Principles

Comparison groups could be added to any of the designs shown in Figure 2-5. If a comparison group was added to the third design shown, then the design would look like this:

- Pre-Program, Post-Program – with Comparison Group

Comparison groups can provide more information about a program’s impact, but comparison group designs are more difficult. Information has to be collected from participants and non-participants. Also, if the participants and non-participants aren’t similar to start with, the comparison is likely to be faulty. This is why it is a problem to compare people who volunteer for a program to those who do not. Volunteers may be more motivated and interested than non-volunteers.

Normed tests are one way to “create” a comparison group. Standardized measures provide information (called “norms”) that show how people typically perform on the measure. For example, developmental screening measures, such as the Ages and Stage Questionnaire (ASQ) and the Denver II-Developmental Assessment, provide solid norms for children of different ages (see Table 6-5). Programs can compare their results with these norms.

NORMS

Can serve as a comparison group for your participants

Norms are available for many of the measures (especially the more complicated ones) reviewed in Chapters 5-9. One great advantage of using normed measures is the ability to contrast how your participants perform on that measure compared to the norms. In essence, the norms are your comparison group. Norms are most useful when your participants are similar to the people on whom the norms were created.

Sometimes, however, programs compare high-risk families outcomes to norms for “average” (not high risk) families. When this is done, remember that average scores for high-risk families may indicate very positive outcomes. For example, the Oregon Healthy Start Program provides intensive home visitation to families at high risk of poor child outcomes, including child maltreatment. In 1996, using normed measures indicate that *among these high risk families*:

- rates of child maltreatment were similar to the general U.S. population,
- rates of reading to their children were higher than for most U.S. families (the HOME measure – Table 5-36) and
- about 90% of children were making normal developmental progress (Table 6-5).

Principle Six: Use What You Learn

The purpose of performance measurement is to improve public and non-profit programs and initiatives. As the third component of the OCCF Systems Change Initiative, performance measurement is integrally tied to comprehensive planning and community implementation (see Figure 1-1, Chapter 1). These close ties make the effective utilization of information from performance measurement possible. Nevertheless, utilizing performance measurement findings is dependent on several factors, including (Worley & Newcomer, 1997):

- Commitment by top leadership – philosophically and with needed resources and technical assistance
- Commitment by leadership, staff, and advocates to “continuous improvement” in which findings are used primarily to strengthen programs and initiatives – not primarily to inform budget cutting
- Clear relevance of performance information to planning and implementation decisions, and
- Effective, communication of findings.

FINDINGS ARE UTILIZED WHEN

- *Leadership is committed*
- *Continuous improvement is emphasized*
- *Information is relevant and timely*
- *Reports are accessible*

Commitment by leadership, a continuous improvement orientation, and clear relevance to decisions, have been discussed earlier. Here we focus on some basic guidelines for effective communication of findings.

Effective reports clarify facts, meanings, interpretations, and make recommendations. To be useful, information has to be timely, focused, and accessible. Most people, from agency staff to legislators, need to review key findings and information quickly. Accessible reports are focused and clearly separate facts, meaning, judgements, and recommendations.

The facts. In creating any report, written or verbal, first report the facts of the assessment.

- Describe the logic model of the program.
- Describe what, how, and from whom information was collected (the design).
- Answer the question: what evidence is there that the program achieved its intended activities and outcomes? Report frequencies, percentages, average scores, or other evidence in summary form.

Principles

The meaning of the information. Next highlight your analysis of the information. Answer the questions:

- Why is this finding important?
- How does this finding related to healthy children, positive youth development, educational success, strong families, or some other wellness goal or Benchmarks?

MEANING

As a result of this a program, parents read to their young children more often each week...such reading is a strong predictor of "readiness for school at age 5."

Judgments. The next step is to state any interpretations or judgments you may make. Judgments involve statements about "success" or other values. Answer the questions:

- Given these findings and their meaning, how do you judge the program?
- Is it successful in reaching its intended outcomes, or not so successful?
- Is it successful for some participants but not others?

Recommendations. Making appropriate recommendations is the final step. Based on the facts, their meaning, and your judgments, what courses of action are possible? Recommendations often include answering questions, such as:

- What changes might be made in the program? Should it stay the same?
- Do the intended outcomes need to be revised ?
- What are the alternative courses of action to improve the effort and outcomes for participants?

In making recommendations, you are speculating about how to improve or strengthen an effort. These are informed recommendations that emphasize continuous improvement and that are *based* on what was learned in the performance measurement process.

RECOMMENDATIONS

Emphasize continuous improvement of program operation and assessment

Recommendations may also include ideas for improving the performance measurement process itself. If there is evidence that the measures failed to capture positive results that staff and participants believe occurred, future efforts should more carefully define outcomes and measures. Remember, continuous improvement applies to performance measurement as well as programs.

NOTE: On the following two pages a short report is presented on the Oregon State University Expanded Foods and Nutrition Program (EFNEP). This report is intended to inform administrators, legislators, and the public about the outcomes of EFNEP. Look for the use of facts, meanings, and judgements.

Some Final Words

Performance measurement in public and non-profit programs is growing. More and more often, this includes assessment of outcomes or results. As performance measurement grows, it is important to remember its six most basic principles.

SIX PRINCIPLES

Know what performance measurement is

Be useful, accurate, feasible, and respectful

Begin with a program logic model

Know your capacity for assessment

Know your design: Who, when, what and how

Use what you learn

Public and non-profit programs are *learning* to effectively operate on-going performance measurement systems. Within the general principles outlined above, experimentation is essential!

EXPERIMENTATION

is essential to develop effective performance measurement.

The most important outcome of performance measurement will be the improvement of programs and initiatives that serve people. It is critical that what is learned from performance measurement be included in decision-making and future planning. Performance measurement

that emphasizes continuous improvement is a great asset to all programs and communities.

In contrast, punitive uses of performance measurement are likely to undermine program improvements. If agencies fear that performance measurement will be used to undercut their support, they may be less willing to serve the hard-to-serve or to seek challenging results.

Emphasize learning and continuous improvement

It will take time, and resources, to plan, implement, try out, adjust, and improve performance measurement systems. It will also take commitment by staff, administrators, decision-makers, legislators, and other stakeholders.

Remember it is worth the effort to do it right!

