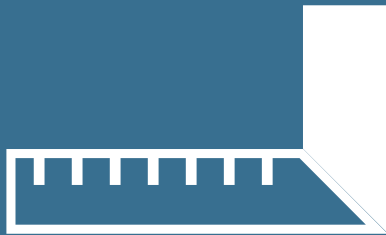




National Child Welfare Resource Center
on Legal and Judicial Issues

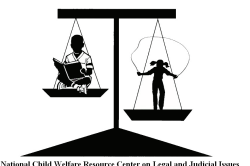


A Guide to Conducting Effective Training Evaluation

Recommendations, Strategies,
and Tools for Dependency Court
Improvement Programs

Sophia I. Gatowski, Ph.D.
Shirley A. Dobbin, Ph.D.
National Council of Juvenile and Family Court Judges

1st Edition, Version 1



The development of this Guide has been funded by the U.S. Department of Health and Human Services Children's Bureau through the National Child Welfare Resource Center on Legal and Judicial Issues (NRCLJI).

ACKNOWLEDGEMENTS

The authors would like to thank the National Child Welfare Resource Center on Legal and Judicial Issues (NRCLJI) and its partnering organizations (the American Bar Association Center on Children and the Law, the National Council of Juvenile and Family Court Judges, and the National Center for State Courts), for support of this publication. NRCLJI is especially grateful to those state court improvement programs (CIPs) that were willing to field-test the strategies and tools contained within this Guide and collaborated with the authors on training evaluation approaches. The insights gained during these technical assistance relationships with CIPs brought unique insights to the project, ensuring its broad coverage and accessibility to the dependency court training context.

Special thanks to Alicia Summers, Ph.D., of the National Council of Juvenile and Family Court Judges who worked closely with the authors on earlier drafts of the Guide. Alicia's careful review, expert contributions to content areas, and valuable feedback on the Guide's direction and messaging, were critical to the completion of this document. The authors would also like to thank Joshua Padilla, Ph.D., and Mimari Hall, M.A., of the National Council of Juvenile and Family Court Judges for the helpful research and input they provided on early drafts of the Guide. Theresa Bohannon, MPH, also of the National Council deserves credit for enhancing the Guide's usability, content, and for working closely with several states to implement the Guide's tools.

And finally, a special thank you to Nancy Miller, former Director of the Permanency Planning for Children Department of the National Council of Juvenile and Family Court Judges, whose constant support, timely encouragement, and substantial guidance given throughout the development of this Guide were most appreciated.

DEDICATION

This guide is dedicated to the late Dr. Shirley Dobbin, who along with Dr. Sophia Gatowski, her dear friend and intellectual partner, wrote this guide. Dr. Dobbin spent her career making training meaningful for judges across the United States. Her boundless contributions to the field will positively impact the lives of children and families for many years to come.

TABLE OF CONTENTS

Introduction.....	1
The Guide to Conducting Effective Training Evaluations.....	2
The State of Dependency Court Training Evaluations.....	3
The Purpose and Goals of the Guide.....	4
Overview of the Guide's Sections and Tools and Resources.....	4
Chapter One: Training Program Planning and Evaluation.....	7
Section 1.1 Training Evaluation Basics.....	8
Section 1.2 Instructional Design Models and Training Evaluation.....	12
Section 1.3 Training Planning and Design Considerations.....	22
Section 1.4 Preparing Adults to Learn.....	26
Section 1.5 Planning Trainings with an Evaluation Focus.....	35
Section 1.6 Designing Measureable Objectives.....	44
Section 1.7 Using Logic Models.....	49
Chapter One: References and Resources.....	52
Chapter Two: Training Satisfaction and Reaction Measurement.....	55
Section 2.1 The Basics of Training Satisfaction and Reaction Measurement.....	56
Section 2.2 The Survey Process.....	65
Section 2.3 Question Construction.....	70
Section 2.4 Online or Web-based Surveys.....	78
Section 2.5 Interviews and Focus Group Methods.....	81
Section 2.6 Survey Sampling.....	85
Chapter Two: References and Resources.....	87
Chapter Three: Measuring Learning Acquisition.....	89
Section 3.1 Learning Acquisition Evaluation Basics.....	90
Section 3.2 Methods for Measuring Learning.....	93
Section 3.3 Guidelines for Accurately Measuring Learning.....	100
Chapter Three: References and Resources.....	104
Chapter Four: The Assessment of Behavior and Practice Change.....	105
Section 4.1 Evaluating Behavior and Practice Change Basics.....	106
Section 4.2 Evaluation Designs for Measuring Behavior and Practice Change.....	108
Section 4.3 Methods for Assessing Behavior and Practice Change.....	112
Section 4.4 Guidelines for the Measurement of Behavior and Practice Change.....	117
Chapter Four: References and Resources.....	118
Chapter Five: The Assessment of Training Outcomes.....	119
Section 5.1 Evaluating Training Outcome Basics.....	120
Section 5.2 Strategies for Facilitating Outcome Measurement Evaluation.....	125
Section 5.3 Major Models of Outcome Assessment.....	131
Section 5.4 Data Collection Methods for Assessing Training Outcomes.....	137
Chapter Five: References and Resources.....	139
Chapter Six: Analyzing, Interpreting and Reporting Training Evaluation Findings.....	140
Section 6.1 Analyzing and Interpreting Evaluation Data.....	141
Section 6.2 Reporting Training Evaluation Findings.....	151
Chapter Six: References and Resources.....	155

INTRODUCTION

Each year millions of federal, state, and foundation dollars fund local training initiatives designed to make a difference in the lives of children and families. Practitioners concerned with the delivery of small- and large-scale training initiatives devote countless hours of their time to the nuts and bolts of program delivery. Practitioners provide direct services to participants, facilitate discussions with program partners, identify faculty members, and address unanticipated training-related issues as they arise. Meanwhile, uneasy thoughts often lurk somewhere in the back of their minds – *Are training goals being achieved? How will that be known? How will evaluation occur?* The evaluation of training programs can seem daunting. Conducting a training program evaluation is often viewed as a waste of precious time, energy, and other resources when resources are scarce. And yet, inevitably, evaluation questions start surfacing – *Are the participants benefiting from the training? How can the training program be improved and strengthened? How can this training program demonstrate that it should be re-funded? Is there an impact on policies, programs, and increased knowledge, which ultimately affects the outcomes of children, parents, and families? Is training making a difference?*

Developing and implementing effective dependency court training programs is challenging. With the expanding role of the court in dependency cases, there is an increasing need for training programs that: effectively convey changes in federal and state law; explain and reinforce foundational court best practices and the expanding role of the judiciary, both on- and off-the-bench; and present emerging knowledge in a wide range of disciplinary areas while tying that knowledge to the role of the judge, the practice of the court, and the child and family outcomes achieved. Given increasingly complex training topics, the diverse faculty needed to teach in these areas, and the emerging variety of teaching tools, more attention needs to be paid to how we measure the success and outcome of all types of training programs.

Why do we need to conduct more effective (or informative?) evaluations of training programs?

We need to better evaluate training programs to learn how to better target, deliver, and design trainings. Having evaluation data makes a better argument for ongoing training resources. Without access to evaluation data, dependency court training will never be able to move beyond standard training programs to address complex issues, roles, practices, and theories. We need training evaluations designed to focus on individual practice improvement, system improvement, and ultimately, better outcomes for children and families.

The Guide to Conducting Effective Training Evaluations

With funding from the U.S. Department of Health and Human Services, Children's Bureau, the Permanency Planning for Children Department (PPCD) of the National Council of Juvenile and Family Court Judges (NCJFCJ), in collaboration with the American Bar Association Center on Children and the Law (ABA) and the National Center for State Courts (NCSC), developed *A Guide to Conducting Effective Training Evaluations: Recommendations, Strategies and Tools for Dependency Court Improvement Programs* (Guide). The partners (NCJFCJ, ABA, and NCSC) recognized a need to help training coordinators better evaluate their training programs, including assessing training outcomes, and explaining how to use evaluation data to better target, design, and deliver trainings.

The Guide is designed to provide strategies and evaluation tools to training coordinators about how to identify training needs, develop training methodologies, assess training outcomes, and develop and implement training evaluation tools. This Guide does not provide "standard" training evaluation forms or a "standard" model to apply to training programs. The goal of the Guide is not to suggest a "cookie-cutter," or "one-method-fits-all" approach to training evaluation. While template evaluation questions are offered, the intent is for training managers to use those templates as a jumping-off point for the development of their own training evaluation designs and instrumentation. Also, the Guide is not meant to be a comprehensive primer on curriculum design, implementation, and evaluation research methods. Rather, the Guide provides an overview of the key steps in designing and implementing an effective training evaluation, and offers recommendations and strategies for determining if your training program successfully met its objectives and resulted in desired outcomes.

The information and tools contained in the Guide have been gathered from a comprehensive review of literature on effective strategies for adult education and training programs, an extensive review of evaluation resource materials, including existing evaluations of dependency court trainings, as well as interviews with judicial educators and Court Improvement Project (CIP) Directors. We examined options for training design and mode of delivery, and obtained perspectives on the state of training evaluation generally and for dependency court systems specifically – what could and should be improved and how? An Advisory Committee of CIP Directors, judicial educators, and evaluation experts was convened to give recommendations on the Guide's

approach and content. In addition, a number of state CIPs participated in field-tests of the strategies, guidance, and template evaluation tools offered in the Guide. Feedback and lessons learned from these field-tests were used to enhance the Guide's final materials and recommendations.

The State of Dependency Court Training Evaluation

With respect to CIP and dependency court trainings, the individuals interviewed reported significant challenges and constraints to conducting the kind of training evaluation they felt was truly needed. Many challenges were noted by participants, including lack of funding for evaluation activities, lack of staff to devote to evaluation activities, lack of value placed on evaluation, failure to plan for evaluation from the onset of program planning and design, and significant time constraints. Our review of evaluation materials also found that most dependency court training evaluations conducted have been limited to exit surveys based on a few general questions (typically administered via a questionnaire at the conclusion of a training session or program). These surveys have primarily resulted in evaluation reports of numbers of people trained, professional groups trained, reactions to the training (e.g., content, faculty, venue), and satisfaction with the training experience.

Our review also discovered that the training outcomes of change in behavior and attitude were rarely measured; a finding that interviewees also reiterated. If change in behavior and attitude were measured, they were mostly assessed through self-report (e.g., a question on an exit survey asking participants if they intended to implement what they learned after the training or if the training had changed their thoughts and attitudes about an issue or problem). Beyond behavior and attitude change, training impacts and outcomes were infrequently measured. For example, there were very few training evaluations that incorporated the research designs necessary to begin to associate training programs with specific impacts. This lack of impact or outcome evaluation makes generalizability problematic. Without impact data, we cannot know what specific training modules or components should be applied to specific training issues in order to have desired results. Without knowledge of the impact or outcomes of training, you cannot know if replicating that specific training model in your jurisdiction, or in other jurisdictions, will achieve intended results.

In addition, our review of dependency court training evaluation materials found significant measurement problems, making valid evaluation difficult, if not impossible. For example, there was a lack of specificity about training goals and learning objectives, a lack of specificity about desired training outcomes, a lack of precision in evaluation instrumentation, and low response rates (i.e., few participants actually completed and returned

Challenges to Training Evaluations

- Lack of funding
- Lack of staff to devote to evaluation activities
- Lack of value placed on evaluation
- Failure to plan for evaluation from onset of program planning and design
- Time constraints

Problems with Current Training Evaluations

- Behavior and attitude change rarely measured
- Training impacts and outcomes rarely measured
- Significant measurement issues

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

training evaluation materials). Using multiple data collection methods, such as combining surveys with interviews, behavior and practice observations can provide a more robust examination of whether training goals were met. However, most training evaluations failed to use more than one method for obtaining data, typically focusing on a questionnaire or survey to obtain evaluation findings.

The Purpose and Goals of the Guide

This Guide is not meant to serve as a textbook for designing and implementing a training evaluation. Rather, the purpose of this Guide is to serve as a user-friendly resource that will assist State CIPs, NCJFCJ Model Courts, and individuals tasked with the design of dependency court training across the nation to design and implement effective training evaluations – to use methods that allow training coordinators to go beyond reporting the number of training programs held and the number of participants trained; and to assess the impacts and outcomes that can be attributed to the training. The Guide does not prescribe an approach or method for training evaluation, but instead offers template tools and recommends evaluation strategies that are intended to be adapted to specific training programs. Feedback provided by reviewers on earlier drafts of this Guide highlighted the need to focus on training program design, stressing the importance of planning for evaluation at the early stages of training development and implementation.

The goals of the Guide are to:

- Provide guidance and strategies for the program planning, design, implementation, and evaluation stages of dependency court training efforts;
- Encourage training coordinators to use the strategies offered in this Guide to support local, state, and national training agendas; and
- Provide template tools that facilitate an assessment not only of satisfaction but also of skill or knowledge acquisition; behavior, practice and attitude change; and training impacts or outcomes.

Overview of the Guide's Sections and Tools and Resources

Chapter One: Training Program Planning and Evaluation. This Chapter of the Guide focuses on the critical program planning tasks, which support effective trainings, including the need to plan for evaluation from the onset of training program design. Chapter One sections cover training evaluation basics; the four levels of evaluation measurement; using internal or external evaluators; training program design considerations, including preparing adults to learn; developing measureable objectives; and using logic models as a tool for evaluation as they help articulate the training theory and expected outcomes. The Tools and Resources accompanying this Chapter includes: 1) training program planning tools; 2) worksheets to help in planning evaluations; 3) tools for designing instructional methods; 4) strategies for conducting effective training needs assessments; 5) tools for designing measureable objectives; and 6) logic modeling worksheets.

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

Chapter Two: Training Satisfaction and Reaction Measurement. This Chapter of the Guide provides recommendations and sample strategies for ensuring that the measurement of training participants' satisfaction is precise, valid, and provides useful information to training organizers. Chapter Two sections include the survey process; good questionnaire construction; designing measurement or response categories; increasing response rates; and the use of online measures. These topics are relevant and foundational to each of the methodologically-focused Sections of the Guide (see below) and should be referred to again when considering learning acquisition, behavior change, and outcome assessment. The Tools and Resources accompanying this Chapter includes: 1) sample satisfaction measurement questions; 2) sample satisfaction and reaction evaluation forms; 3) interview and focus group protocols; 4) strategies for designing response categories; and 5) considerations when using online or web-based surveys. In addition, a Question Bank is provided with the Tools and Resources materials, which includes sample training satisfaction and reaction measurement questions, and response formats for the reader to adapt to their own training evaluation needs and context.

Chapter Three: Measuring Learning Acquisition. This Chapter of the Guide focuses on measuring learning. Chapter Three sections cover the primary types of learning that can occur during training programs and the methodologies evaluators can use to properly measure learning acquisition. The Tools and Resources accompanying this Chapter includes: 1) templates for measuring learning; 2) pre- and post- training learning assessment forms; 3) sample protocol for testing learning acquisition; and 4) creative approaches to measuring learning. The Question Bank provided with the Tools and Resources materials includes sample questions designed to tap learning acquisition.

Chapter Four: The Assessment of Behavior and Practice Change. An important goal of training is to ensure that behavior or practice changes occur as a result of the training – that training participants apply what they learned at the training. This Chapter provides guidance regarding how to assess whether trainees applied their newly acquired knowledge or skills, or modified their attitudes and behaviors, in the performance of their role, job, or duties. Chapter Four sections cover the importance of instructional designs that support behavior and practice change, evaluation designs and methods for assessing behavior and practice changes, and guidelines for developing and implementing training evaluations to assess behavior and practice change. The Tools and Resources accompanying this Chapter includes: 1) a sample behavioral self-assessment form; 2) sample observational tools to measure behavior and practice change, including court observation and supervisor/mentor behavioral assessment tools; and 3) information on using secondary sources to assess behavior and practice change such as case file or document review. The Question Bank included with the Tools and Resources also provides sample questions designed to measure self-report of behavior and practice change.

Chapter Five: The Assessment of Training Outcomes. This Chapter of the Guide reviews approaches to determining the outcomes associated with a training program. Chapter Five sections build upon the principles and guidance offered in each of the previous Chapters, including defining outcome measurement and

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

associated terminology; planning for outcome evaluation at the training design phase; defining relevant outcomes; major models for outcome evaluation; and data collection methods for measuring outcomes. The Tools and Resources accompanying this Chapter includes: 1) key steps to implement an outcome-focused training evaluation; 2) examples of outcomes, indicators, and data sources for outcome measurement; and 3) strategies for identifying outcomes.

Chapter Six: Analyzing, Interpreting, and Reporting Training Evaluation Data. The final Chapter of the Guide provides suggestions for analyzing, reporting, and using your training evaluation data so that all of the hard work put into the evaluation will be meaningful for training program improvement. Chapter Six sections cover analytic models or frameworks to apply to understand your evaluation findings and how to report findings in ways that are easily understood. The Tools and Resources accompanying this Chapter includes: 1) additional analytic tip sheets; and 2) sample report outlines.

As mentioned above, a series of companion documents to this Guide are the **Tools and Resources**. The Tools and Resources provide sample checklists, worksheets, template evaluation tools, and other resources and recommendations to assist in the design and implementation of effective training evaluations. Look for the measuring tape symbol through the text of this Guide for relevant tools contained in the Tools and Resources for that Chapter.

CHAPTER ONE

Training Program Planning and Evaluation

This Chapter of the Guide emphasizes that the development, implementation, and evaluation of training programs needs to be an ongoing process (see figure 1) that builds upon past learning opportunities, uses new teaching strategies, engages new and emerging topical areas, and supports ongoing systems' reform efforts. As training organizers are designing their curriculum and preparing for its implementation, it is critically important to plan for the inclusion of an evaluation.

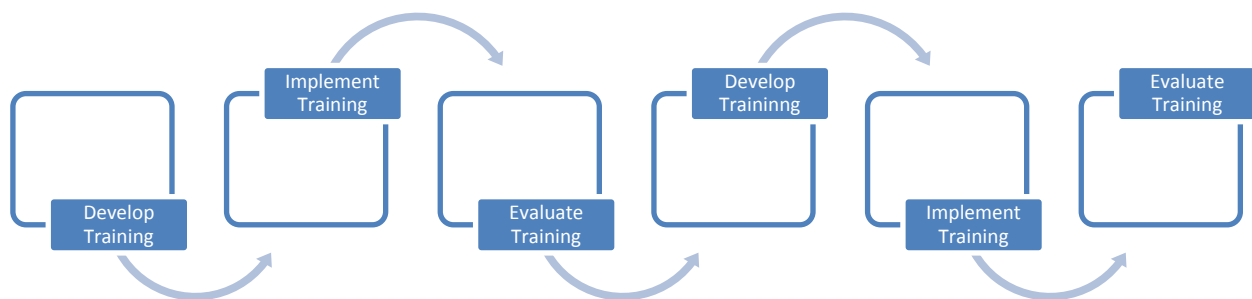


Figure 1

The sections in this Chapter discuss some training evaluation basics; including: information on why training organizers should conduct evaluations (the benefits and risks); basic questions used to determine a training evaluation approach; the domains of training evaluation measurement; the four levels of training evaluations (satisfaction, learning, behavior, results); and how knowing the reason for conducting the evaluation can help organizers determine which level of evaluation is appropriate for their training program. The Chapter concludes with a discussion of specific training program planning considerations and planning tools that will facilitate the development and implementation of effective trainings and training evaluation.

SECTION 1.1

Training Evaluation Basics

Given the increasing complexities that courts, attorneys, child welfare agencies, and a wide range of system professionals are facing to address the needs of abused and neglected children and their families, training and educational needs of must be fully understood before learning how to develop and implement more effective training programs.

Some promising training programs with good evaluation results have been conducted over the years. We need to learn from effective training programs and use what we learn to inform our current and future trainings – what are effective teaching methodologies, effective strategies, effective trainers, etc.? Some poor training programs have also been conducted over the years. Why did certain training programs result in poor or ineffective outcomes? It is no longer sufficient to just report the number of training programs and the number of participants or to rely on standardized training programs without knowledge of their outcomes. The development, implementation, and evaluation of training programs needs to be an ongoing process that builds upon past learning opportunities, utilizes new teaching strategies, identifies new areas of training need, engages new and emerging topical areas, and supports ongoing system reform efforts.

Reasons to Conduct Evaluations of Training Programs

A primary reason training evaluations are conducted is to determine the effectiveness of the training program. In this scenario, evaluations are primarily used to elicit information that will help organizers improve future training programs. Evaluations serve as a tool to refine training programs until they accomplish their intended effect. The newer a training program, the more critical it is to conduct a thorough evaluation. The information elicited usually pertains to the presentation, organization, and substance of the training curriculum. Trainings that incorporate evaluations and apply what they learn are able to improve their presentations over time. As a training program establishes a routine that works, evaluations serve to ensure the training is continuing to meet the goals of trainees.

Another reason training evaluations are conducted is to decide whether to continue or discontinue training programs. Despite best efforts to improve the training program, some training never achieves its desired or intended effect. If evaluations continually indicate that the training has little effect on the trainees, the training is subject to termination. If never evaluated, promising training programs may be cancelled and funding withdrawn. Training evaluations are also used to justify the existence of specific training programs.

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

Whether training is funded privately, or by a non-profit or government entity, current funding requirements often require an evaluation component to prove the utility of the work done. Similarly, evaluation can be used to show how a specific training program contributes to an organization's overall objectives and goals.

REASONS TO CONDUCT EVALUATIONS OF TRAINING PROGRAMS

The benefits of evaluating training programs ...

- Justifies training budgets.
- Expands training budgets.
- Obtains and analyzes user satisfaction feedback.
- Identifies best training strategies.
- Identifies best and most appropriate content areas.
- Relates training to needed practice areas and outcomes.
- Clarifies training expectations for professional activities.
- Provides an opportunity to assess professional development.
- Ties training objectives to organizational and system objectives.
- Quantifies performance improvement of individuals, institutions, programs, and child-based outcomes.

The risks of not evaluating training programs ...

- Wastes dollars on ineffective programs.
- No method to assess training dollars spent.
- Discontinues training program.
- Inefficient and ineffective delivery of training.
- Training program does not meet objectives.
- Users' training needs not met or remain unknown.
- Performance improvement unknown.
- Difficult to assess professional development.
- Performance expectations are not set or communicated.
- Unknown if and how training program impacts practice, policy, and law.
- Unknown if training program improves system outcomes, especially with respect to child-based outcomes.

Planning a Training Evaluation

While the reasons for conducting a training evaluation may be clear, there are a series of questions that can be used to help training organizers gauge the level of evaluation that will be required, which will help to determine the evaluation domain and level of measurement.

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

The following questions should be addressed when planning a training program and its evaluation:

- *What do I want to know?* What are the intended goals of the training program? What are the learning objectives of the training program? What do we hope to achieve? Training evaluations should be designed to measure whether the training has met its goals, learning objectives, and intended outcomes.
- *What is the best mode of evaluation?* Organizers have to determine the best way to collect the information they are seeking. Which of the data collection methods (e.g., paper or online survey, interview or field-observation) lend themselves to the resources (staff, time, and budget) available? Here again it is important to consider the intended learning objectives for the training.
- *What is the timeframe for asking questions?* Organizers have to determine the optimal time to conduct a training evaluation. Organizers may implement a series of data collection methods at various stages of the training, such as paper and pencil or an online survey conducted before and immediately after trainings, or in-depth interviews that can occur up to a few weeks after the training has concluded. Evaluators may also want to conduct a fidelity check months after a training to ensure trainees have incorporated the behaviors they learned into practice (later Sections of the Guide provide more information about fidelity checks). The appropriate timeframe for any given training depends on the specific learning objectives for the training.
- *What do I intend to do with the information received from the evaluations?* After determining the optimal mode of collecting information, organizers must decide how they want to use the information. Organizers can vary what they do with information from keeping it in-house to developing a formal report for dissemination or publication. Knowing what you intend to do with the information will direct the number and type of questions you ask in your evaluation.

A systematic approach to evaluation planning is organized around the following 10 questions:

1. What is being evaluated?
2. Why is the evaluation being conducted?
3. What are the main issues/ questions to be addressed by the evaluation?
4. Who will do what?
5. What are the resources for the evaluation?
6. What data needs to be collected?
7. How will data be collected?
8. How will the data be analyzed?
9. What will the reporting procedure be?
10. How will findings from the evaluation be used?

Answering evaluation planning questions outlined in the text above allows training organizers to determine the purpose and scope of their evaluation. Training evaluations should always begin by thinking about the

information they are seeking, the best time and method for collecting it, and what you intend to do with the information you ascertain. Understanding the reasons you are conducting an evaluation and considering what you intend to do with the information you collect is the first step in conducting an evaluation. It allows you to narrow your focus into the specific components necessary to conduct a successful evaluation.

Domains of Training Evaluation Measurement

One of the fundamental purposes of formally evaluating a training program is to determine if the training program was efficient and effective, and whether or not it achieved the appropriate level of impact.¹

- **Training Efficiency** – an examination of the cost of the training program. Was it worth it?
- **Training Effectiveness** – an examination of whether the training program resulted in substantive changes in knowledge and skills in each training participant. The effectiveness of faculty and trainers, as well as the effectiveness of training strategies, needs to be determined.
- **Training Impact** – an examination of the impact of the training on participants' attitudes and behaviors, as well as changes in practice, policy, and procedures. Ultimately, an examination of whether the training program is appropriately impacting achieved outcomes for children and families.



Figure 2

Evaluations need to focus on training efficiency, training effectiveness, and training impact, as seen in figure 2. We need to build upon what works and learn from what does not work. We need to be able to demonstrate the effectiveness and importance of training programs to ensure support for further training. We need to broaden areas that support individual and system-wide learning and practice change. We need to ensure that training programs are effectively moving courts, and the broader system, to the achievement of improved outcomes for abused and neglected children.

¹ Boulmetis, J. & Dutwin, P. (2005).

SECTION 1.2

Instructional Design Models and Training Evaluation

Several strategies can be applied to evaluate a training program. The strategy selected depends upon what you want to know from the evaluation – your specific learning objectives and desired or intended outcomes are critical to the selection of your evaluation approach. After determining the reason for conducting an evaluation, organizers must determine the level of evaluation necessary to determine if their goals will be met.

There are a number of evaluation models, frameworks, or approaches that can be applied to a training evaluation. The purpose of this Section is not to go through these in detail, but instead to provide readers with some exposure to a useful tool—the instructional design model—that can be used in determining an evaluation approach. Instructional design models provide training organizers with the destination for the training (its goals and objectives) and the road map to get there (the paths the instruction will take to facilitate learning).

Instructional design is a tested and proven methodology for developing instruction or curricula, and refers to having a formal model for training design, implementation, and evaluation. Instructional design is a tool that helps you create an effective training program in an efficient manner. It is a framework that helps you ask the right questions, make the right decisions, and ultimately produce a training product that will be useful for your training audience. Creating a training program without an instructional design model in mind (or specified theory of learning) may make it more difficult to determine the impact of your training program effectiveness. Training programs that are implemented without an instructional design model may lead training organizers to focus evaluations solely on whether trainees “liked” the program and not on whether learning actually took place.

While a comprehensive instructional design approach may not always be necessary (see the Section One Tools and Resources for a checklist to help you determine whether an instructional design approach is recommended), instructional design can be used to create a valid and useful evaluation of the training. Through the use of instructional design procedures, for example, you design objectives for the training program that can be used as the basis of evaluation, determining which objectives your trainees have met. Because instructional design bases these objectives on an assessment of the needs or competencies of trainees, you can relate the training to real world applications and determine if trainees are actually using what they have learned.

In an instructional design approach to training planning, implementation, and evaluation:

- Rigorous analysis identifies the most critical content material and makes it a priority of the training program.
- The training program is tailored to a specific audience.
- Training content is derived from learning objectives that specifically state what individuals are to learn.
- Training content is delivered in a variety of methods and carefully planned to evoke maximum learning and retention.
- Evaluation is varied and frequent.
- Evaluations are tied to learning objectives.
- Evaluations are used to provide outcome data (feedback about the overall effectiveness of the training) and a tool for training organizers to constantly improve the delivery of training program.

While the list of instructional design models provided below is by no means comprehensive of all of the instructional design models that have been recommended, a few of the most commonly applied and practiced instructional design models, theories, and methodologies are summarized.

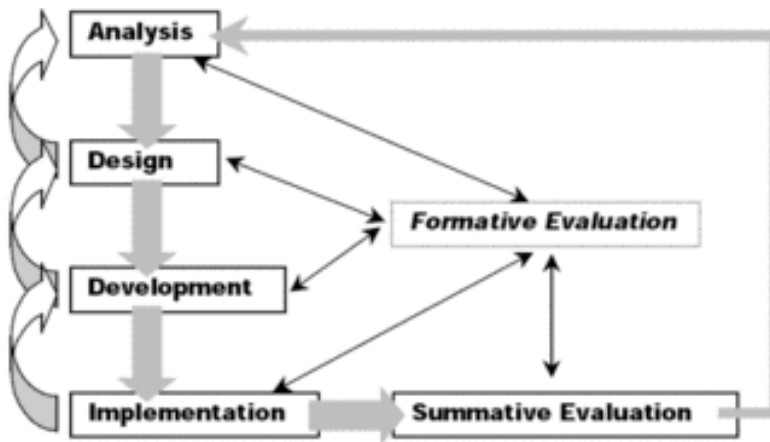
A Sampling of Instructional Design Models

ADDIE Instructional Design Model (Analysis, Design, Development, Implementation, Evaluation)

The ADDIE instructional design model² outlines a generic process traditionally used by instructional designers and training developers. The five phases—Analysis, Design, Development, Implementation, and Evaluation—represent a dynamic, flexible guideline for building effective training programs. As demonstrated in figure 3, each step in the ADDIE model has an outcome that feeds into the next step in the sequence.

- Analysis – in this phase, the instructional problem or training need is clarified, the instructional goals and objectives are established, the learning environment identified, and the learner's existing knowledge base and skills are determined.
- Design – in this phase, learning objectives for content areas are fully articulated; assessment instruments, exercises, content, subject matter and faculty are identified; and the mode of training delivery is selected.
- Development – in this phase, training organizers develop and assemble the content and instruments that were identified in the design phase (this is the curriculum writing and curriculum finalization phase).
- Implementation – in this phase, the training is implemented.
- Evaluation – in this phase, the training is evaluated (both formative and summative evaluation).

² Dick & Cary (1996); Leshin, Pollock & Reigeluth (1992).



Formative Evaluation – typically conducted during the development of a program, with results being fed back into program procedures for improvement purposes

Summative Evaluation – provides information on the program's effectiveness

Figure 3
ADDIE Model, Diagram by: Steven J. McGriff, Instructional Systems, College of Education, Penn State University

Kemp's Instructional Design Model

The Jerrold Kemp (1985) instructional design method and model defines nine different components of an instructional design and at the same time adopts a continuous implementation and evaluation model. In this model, the design and development process is a continuous cycle that requires constant planning, design, development, and evaluation to insure effective instruction, which is shown in figure 4. Kemp (1985) identifies nine key elements:

1. Identify instructional problems, and specify goals for designing an instructional program.
2. Examine learner characteristics that should receive attention during the training's planning.
3. Identify subject content, and analyze task components related to stated goals and purposes.
4. State instructional objectives for the learner.
5. Sequence content within each instructional unit for logical learning.
6. Design instructional strategies so that each learner, with different learning styles, can master the objectives.
7. Plan the instructional messages (presentations and workshops) and delivery.
8. Develop evaluation instruments to assess objectives.
9. Select resources to support instruction and learning activities.

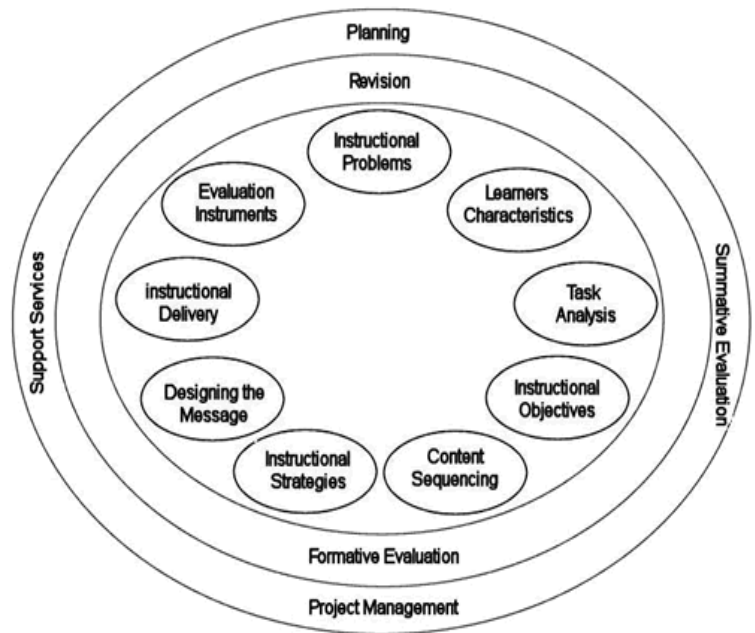
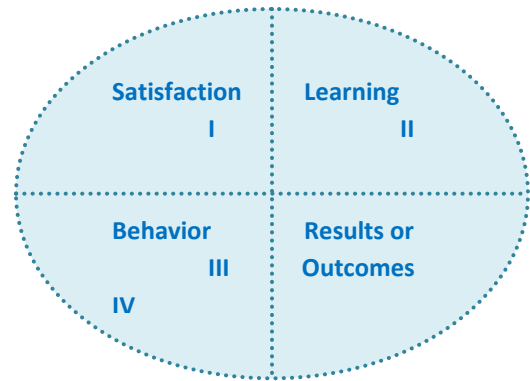


Figure 4
From: Kemp, J.E. (1985). *The Instructional Design Process*. New York: Harper and Row.

The Four Levels of Training Evaluation

Arguably, the most widely used and popular model for the evaluation of training and learning is Donald Kirkpatrick's four levels of evaluation: satisfaction or reaction, learning, behavior, and results or outcomes.³ Each level requires differing degrees of measurement and analysis, resulting in varying levels of information. These concepts are briefly introduced here, and form an organizing framework for the remaining sections of the Guide.



- I. **Satisfaction** – The initial level of evaluation in Kirkpatrick's model involves measuring participant satisfaction (this is also referred to as reaction measurement). Measuring people's satisfaction provides insight into whether trainees believe the training was useful to them. These inquiries typically involve whether participants like the format, the presenter, the facilities, and/or the focus of the information. While satisfaction may be the most basic level of evaluation, it is a key feature that needs to be evaluated to ensure the consumer of the training was generally receptive of the training. Measuring satisfaction typically involves a short list of questions, usually answered as yes/no, in surveys or interviews that occur directly after the training. Participants are usually receptive to this process, and do not mind providing their opinions regarding their general perceptions of the training.
- II. **Learning** – The second level of evaluation involves measuring learning. The goal of many training programs is to instill a certain level of knowledge or skills in training participants. Measuring learning involves providing measures that examine the knowledge and skills participants' acquire from the training. The primary method to accomplish this is to develop questions around the content and substance of the material that was presented. Learning assessment is considered the "second level" of measurement because of the extra considerations required to measure learning or knowledge acquisition. Evaluators must measure participants' level of knowledge and skill both before and after receiving the training. This is accomplished by conducting a pre-test measuring trainees' knowledge level on the topic before the training followed by a post-test measuring trainees' knowledge level on the topic after the program. This process provides training organizers with the ability to determine the effectiveness of the training program in increasing trainees' knowledge.
- III. **Behavior** – The third level of evaluation involves changes in behavior. Examining behavior requires even more detailed measurement than measuring learning. One way to measure behavior change is through self-report; simply asking training participants to report whether their behavior has changed as a result of the training. However, this approach is not ideal. Normal psychological phenomena leave people unable to accurately assess their behavior. For example, research shows that if you ask individuals if they have

³ See for example Kirkpatrick, D.L. (1975).

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

changed their behavior, they will typically affirm that their behavior has changed, even when it is not entirely true.⁴ In order to truly assess whether behavior has changed as a result of training, meticulous care must be taken to measure behavior *prior* to the training. To objectively know whether behavior has changed, you will need an unbiased examination prior to the training to establish a baseline of behavior. It also requires fidelity checks to ensure that models of practice taught at the training are implemented in the field as intended. In addition, post-training measures that are taken days, weeks, or even months after the program are used to determine if changes are immediate or long-lasting.

- IV. Results or Outcomes** – The final level of training involves examining all of the potential results or outcomes. This is the most complete of all the levels of evaluation measurement because it incorporates all of the earlier levels to inform an overall evaluation. Organizers can determine: if participants are satisfied with the training they received; if trainees' have truly learned something new; whether trainees' are incorporating it into their job performance; and overall impacts or outcomes. True results or outcome measurements requires quasi-experimental design (comparison of data on an outcome of interest before you implement the training, for example, followed by rigorous observation and quantification of behavior and practice changes after the training has ended). Fidelity checks are also needed at this level of measurement to ensure that the training content is being implemented as intended – are people doing what they were trained to do? When measuring the results or outcomes associated with training, we are moving beyond merely determining whether training has given its participants skills or knowledge. Instead, in results-oriented training evaluation measurement, the focus is on whether or not the training has impacted a defined issue or problem. Training has been defined as an intervention; a specific strategy or intervention with a participant that is designed to impact behavior, practice, the field, or a social or organizational problem.

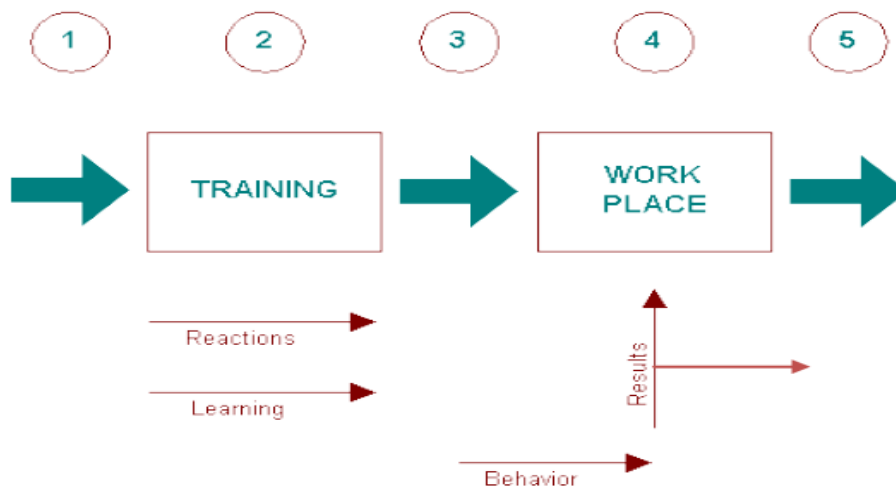
Which one of the four levels of evaluation to be employed is influenced by the goals of the training, the research resources available, and the content or substantive focus of the curriculum? If the goal of training is to get people interested in a general topic, a satisfaction evaluation may be most appropriate. If the goal is to introduce new information to trainees, the evaluation must examine at least the second level (learning) to determine whether it is successfully generating knowledge gains. If the goal of training is to change behavior, evaluators must apply the third level (behavior) of evaluation in order to assess the effectiveness of the training at changing desired behaviors. The final stage (results or outcomes) provides a full examination of every aspect of a training program, and is applied to trainings that are interested in associating specific impacts with the delivery of a training program; moving beyond determining satisfaction with the training program and whether knowledge has been gained or behavior changed, to assessing a training program's impact on a specific issue or problem.

Since the goals of CIP trainings is almost always gaining knowledge, developing skills, implementing what is learned, and ultimately changing behaviors and practice, CIP training evaluations should attempt to include all four levels of evaluation.

⁴ See for example Kahneman, D., Slovic, P. & Tversky, A. (1982); Sternberg, R.J., Roediger, H.L., & Halpern, D.F. (2007).

Connecting Training to the Context of the Workplace or Field-Setting

In the dependency court context, training typically takes place in support of skill or knowledge development for application in the field (i.e., for application in dependency court systems, with the goal of improving or changing behavior and practice). The relationship between training and the workplace is illustrated in figure 5 below, which depicts five basic points at which an evaluator of training might take measurements, conduct assessments, or reach judgments: 1) before the training; 2) during the training; 3) after training; 4) in the workplace or field-setting; 5) upon exiting the workplace or field-setting.



The four elements of Kirkpatrick's framework are defined below:

1. **Reactions** – Reactions may best be defined as how well the trainees liked a particular training program. Reactions are typically measured at the end of training – at Point 3 in the diagram above. In that case, the evaluation is a “summative” or end-of-course assessment. Reactions can also be measured during the training, even if only informally (e.g., an instructor or observer notes their perceptions of the participants’ reactions). Reaction is also discussed in terms of “satisfaction.”
2. **Learning** – What principles, facts, and techniques were understood and absorbed by the training participants? What the trainees know, or can do, can be measured during and at the end of training but, in order to say that this knowledge or skill resulted from the training, the trainees' baseline knowledge or skills levels must also be known or measured. Evaluating learning, then, requires measurements at Points 1, 2 and 3 – before, during and after training.
3. **Behavior** – In the dependency court context, a training goal may be to change on-the-job or in-the-field behavior. Clearly, any evaluation of changes in on-the-job behavior must occur in the workplace setting itself – at Point 4 in the diagram. It should be kept in mind, however, that behavior changes are acquired in training and they then transfer (or don't transfer) to the workplace or field-setting. As a result, it is useful to assess behavior changes at the end of training *and* in the workplace or field-setting.

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

4. **Results or Outcomes** – This level of measurement is the overall impact of the training (e.g., reduction of costs, increase in quality, increase in quantity, improved morale, increased collaboration, shortened timeframes to permanency, improved well-being, etc.). These factors are also measurable **in the** workplace or field-setting – at Point 4 in the diagram – and beyond the workplace or field-setting – at Point 5 in the diagram.

The following tables illustrate Kirkpatrick's structure and levels of training evaluation in more detail, including examples of tools and possible methods.

Table 1-1			
Evaluation Level and Type	Evaluation Description	Evaluation Tools and Methods	Relevance and Practicability
1. Reaction (Satisfaction)	<p>Reaction evaluation is how the participants felt, and their personal reactions to the training or learning experience, for example:</p> <p>Did they consider the training relevant?</p> <p>Was it a good use of their time?</p> <p>Did they like the venue, the style, timing, materials, etc.?</p> <p>Perceived practicability and potential for applying the learning.</p>	<p>Feedback forms based on subjective personal reaction to the training experience.</p> <p>May include verbal reactions, which can be noted and analyzed.</p> <p>Post-training surveys or questionnaires.</p> <p>Online evaluation or grading by participants.</p> <p>Subsequent verbal or written reports provided by participants to their managers and/or training organizers.</p>	<p>Can be done immediately after the training ends.</p> <p>Relatively easy to obtain reaction feedback.</p> <p>Feedback is not expensive to gather or to analyze for groups.</p> <p>Important to know that people were not upset or disappointed by the training.</p>

Table 1-2			
Evaluation Level and Type	Evaluation Description	Evaluation Tools and Methods	Relevance and Practicability
2. Learning	<p>Learning evaluation is the measurement of the increase in knowledge or intellectual capability from before to after the learning experience:</p> <p>Did the trainees learn what was intended to be taught?</p> <p>Did the trainee experience what was intended for them?</p>	<p>Typically assessments or tests before and after the training.</p> <p>Interview or observation can be used before and after, although this is time-consuming and can be inconsistent.</p> <p>Methods of assessment need to be closely related to the aims of the learning.</p>	<p>Relatively simple to set up, but more investment and thought required than reaction evaluation.</p> <p>Highly relevant and clear-cut for certain training such as quantifiable or technical skills.</p> <p>Less easy for more complex learning such as attitudinal, leadership, or collaborative</p>

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

	What is the extent of advancement or change in the trainees after the training, and was it in the direction or area that was intended?	<p>Measurement and analysis is possible and easy on a group scale.</p> <p>Reliable, clear scoring and measurements need to be established, so as to limit the risk of inconsistent assessment.</p> <p>Hard-copy, electronic, online or interview style assessments are all possible.</p>	development.
--	--	--	--------------

Table 1-3			
Evaluation Level and Type	Evaluation Description	Evaluation Tools and Methods	Relevance and Practicability
3. Behavior	<p>Behavior evaluation is the extent to which the trainees applied the learning and changed their behavior; this can occur immediately and several months after the training, depending on the situation:</p> <p>Did the trainees put their learning into effect when back on the job?</p> <p>Were the relevant skills and knowledge used?</p> <p>Was there noticeable and measurable change in the activity and performance of the trainees when back in their roles?</p> <p>Was the change in behavior and new level of knowledge sustained?</p> <p>Would the trainee be able to transfer their learning to another person?</p> <p>Is the trainee aware of their</p>	<p>Observation and interview over time are required to assess change, relevance of change, and sustainability of change.</p> <p>Arbitrary snapshot assessments are not reliable because people change in different ways at different times.</p> <p>Assessments need to be subtle and ongoing, and then transferred to a suitable analysis tool.</p> <p>Assessments need to be designed to reduce subjective judgment of the observer or interviewer, which is a variable or confounding factor that can affect reliability and consistency of measurements.</p> <p>The opinion of the trainee, which is a relevant indicator, is also subjective and unreliable, and so needs to be measured in a consistent defined way.</p> <p>360-degree feedback (feedback about behavior is provided by self, colleagues, supervisors, clients) is useful</p>	<p>Measurement of behavior change is less easy to quantify and interpret than reaction and learning evaluation.</p> <p>Simple quick response systems unlikely to be adequate.</p> <p>Cooperation and skill of observers are important factors, and difficult to control.</p> <p>Evaluation of implementation and application is an extremely important assessment; there is little point in a good reaction and good increase in capability if nothing changes back in the job or field.</p> <p>Behavior change evaluation is possible given good support and involvement from trainees, so it is helpful to involve them from the start, and to identify benefits for them, which links to level 4 evaluations below.</p>

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

Behavior (Cont'd)	change in behavior, knowledge, skill level?	<p>method and need not be used before training, because respondents can make a judgment as to change after training, and this can be analyzed for groups of respondents and trainees.</p> <p>Assessments can be designed around relevant performance scenarios, and specific key performance indicators or criteria.</p> <p>Online and electronic assessments are more difficult to incorporate.</p> <p>Self-assessment can be useful, using carefully designed criteria and measurements.</p>	
--------------------------	---	--	--

Table 1-4

Evaluation Level and Type	Evaluation Description	Evaluation Tools and Methods	Relevance and Practicability
4. Results or Outcomes	Results or outcome evaluation is the effect on the system, organization or environment resulting from the improved performance of the trainee. Measures would typically be business or organizational key performance indicators, such as: Values, percentages, timescales, and other quantifiable aspects of organizational performance, for instance; time to permanency, rates of reunification; achievement of recognized performance standards; percentage of cases in which foster parents appear at hearings, etc.	<p>It is possible that many of these measures are already in place via normal management information systems reporting and performance measurement protocols.</p> <p>The challenge is to identify which measures are available and how they relate to the trainee's role or input and influence on those measures (and to the learning objectives of the training).</p>	<p>Individually, results of the evaluation are not particularly difficult; but it becomes much more challenging when examining complex organizational or systems' goals.</p> <p>External factors greatly affect organizational performance, which cloud the true cause of good or poor results.</p> <p>Failure to link training content to the trainee's actual role in achievement of desired results will greatly reduce the ease by which any results can be attributed to a specific training.</p>

Building a Focus on Results or Outcome Analysis in Training Evaluation

While Kirkpatrick's Model has been very helpful to the design of training evaluations, its linear outline (Level 1: Satisfaction, Level 2: Learning Acquisition, Level 3: Behavior Change, and Level 4: Results and Outcome Analysis) can lead to a focus on Levels 1 and 2 and not move beyond those levels to measure whether any changes and impacts can be associated with a training's implementation. While Kirkpatrick's model has served us well, it's also created a misconception that Level 1 (satisfaction and reaction measurement) is all one needs for evaluation leading to over-simplified evaluations and minimal analysis of the impact of training at Level 4.

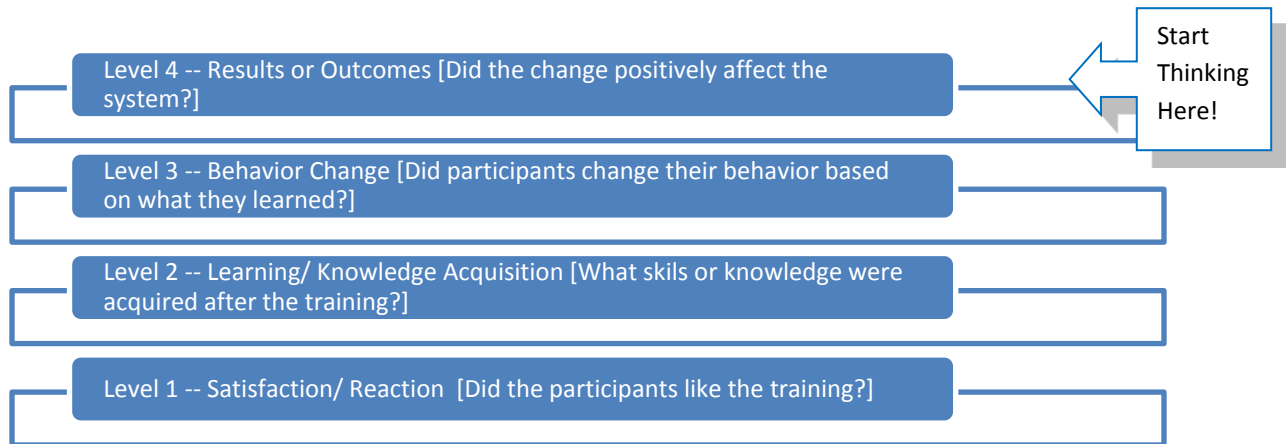


Figure 6

If we were to reverse Kirkpatrick's Model (see figure 6) and turn it on its head, it leads you to think differently about training evaluation. Start at Level 4 – Ask yourself what are the outcomes or results we want from this training? Given those wanted outcomes, what behaviors do you need or want to see from your trainees (a Level 3 analysis)? Given those needed behaviors, what knowledge, skills or attitudes do you want trainees to have (a Level 2 analysis)? Given those knowledge, skills and attitudes, what reactions do we want our trainees to have (a Level 1 analysis)? By reversing the model, the value of Level 4 and Level 3 analysis in terms of providing meaningful information about the effectiveness of your training are highlighted. Reversing your thinking about the model also helps to develop:

- Precise learning outcomes
- Better links among Levels
- Better link of training to outcomes
 - Practically relevant
- Hints for performance assessment/observation
- Better tailoring of training programs
- Better accountability

SECTION 1.3

Training Planning and Fundamental Design Considerations

While Section 1.2 reviewed the value of instructional design to training implementation and evaluation generally, this Section provides some practical strategies for designing and planning trainings. In order for a training program to be effective, you must carefully consider using some form of a training and development plan, even if you are thinking about informal means of training and development. Sketching out a training plan with goals, intended methods, and evaluation approaches, will at least give a sense of what you want and how you will recognize if it has been achieved. Don't expect perfection in the plan or the learner. Start simple, but start. Then update the plan as you go along. Remember your training plan is a general guide; the real benefit found from implementing your plan is the learning you achieve.

While it may seem logical when beginning to design a training to focus on preparing the content of the curriculum, there are some situational issues that if considered simultaneously during curriculum development, can also help to ensure the training is successful and facilitate the evaluation. This Chapter covers basic training design and some of the fundamental issues that should be considered and addressed when designing your training program.

The Training Design Process⁵

Training design refers to the planning and structuring of a training program to achieve specific instructional goals. For trainings to be effective, their features (activities, tasks, subject matter, mode of curriculum delivery, etc.) should be well thought out and designed with attention to desired learning objectives and training goals.

The course design process includes the following activities:

- Identifying appropriate goals
- Choosing content that is consistent with the goals
- Selecting ways to achieve the goals
- Assessing participant learning in relation to the goals

As part of the design process, faculty should also consider:

- Their own teaching style
- The learning styles of the participants
- The role of the course in the overall training effort

Before training begins – Most training design decisions must be made before you step into the training room. These decisions relate to these basic areas:

⁵ Adapted from "Course Design," Center for Instructional Development and Research, The University of Washington, 2004, <http://depts.washington.edu/cidrweb/CourseDesign.html>

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

- The content to include
- The delivery methods to use
- The time allocated for each of the goals
- The tools for assessing participant learning

During the training session – As you conduct the training you will learn more about the participants and their needs. This information may lead you to make adjustments in the course design. For example, after working with the group, you may decide to change the time allocation for a particular topic or change the type of activity associated with a particular topic from an individual to a group activity or vice versa.

At the conclusion of the session – The information you gather at the conclusion of a training session will help you assess the effectiveness of the current training and help you improve future training sessions. To evaluate the course you must use appropriate evaluation tools.

Table 1-5 Some Key Definitions in the Design of a Training Plan			
Learning Objectives	Training Goal	Learning Methods / Activities	Evaluation
What you will be able to do as a result of the learning activities in this plan (e.g., exhibit required skills, a specific practice or behavior).	Overall results or capabilities you hope to attain by implementing your training plan.	What you will do in order to achieve the learning objectives (e.g., implementation of specific curricula modules/sessions, workshops, discussion groups, exercises, etc.).	Assessment and judgment based on evidence in order to determine whether you achieved the learning objectives or not.

Training Planning and Design Tasks

Determine the Needs of Trainees. Each training audience will have unique experiences, knowledge, and occupational practices. Determining the needs of your trainees is fundamental not only to training design but also to training evaluation.⁶ Traditionally, this is accomplished by conducting a needs assessment. Understanding what is needed by trainees is a necessary step in developing effective training programs and evaluation approaches. See Section 1.5 for more detailed guidance on using needs assessments.

Set Achievable Goals. After determining the needs of the trainees, organizers should set specific and achievable goals for the training. Setting realistic goals and measureable objectives is another fundamental factor to be considered when designing a training evaluation.⁷ Resources for training evaluations are usually limited. It is often the case that a training evaluation cannot be designed to meet all of the needs of the trainees in short amounts of time. Therefore, it is important for organizers to set reasonable goals that can be accomplished with the resources available to them. Furthermore, for evaluation purposes, goals should be

⁶ Kirkpatrick, D.L. & Kirkpatrick, J.D. (2006).

⁷ Arnold, W.E. & McClure, L. (1996); Kirkpatrick, D.L. & Kirkpatrick, J.D. (2006).

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

specified in terms of measureable objectives. See Section 1.7 for more detailed guidance on developing learning objectives that are measureable.

Determine the Scope of the Curriculum. Determining the subject content of the training program is critical to preparing for a training evaluation.⁸ The curriculum should be organized in a manner in which the importance and utility of the information are emphasized from the beginning, ensuring buy-in among the trainees. In addition, consideration of learning styles of the audience should also inform the development of the curriculum. Ignoring the principles of adult learning and learning style could lead trainees to become defensive and unmotivated to participate, rendering the training and evaluation ineffective.⁹ See Section 1.4 for more guidance on ensuring curricula incorporate the principles of adult learning and learning style.

Ensure the Training Schedule is Convenient. Scheduling impacts the development of training in two ways. First, it is important to incorporate frequent breaks during training sessions. Adults are typically not able to fully attend to information for more than two hours.¹⁰ The scheduling of alternative activities, lunches, and other breaks is vital to keep trainees' minds fresh and focused. Second, organizers should try to schedule training during the time trainees usually work. If a group of trainees typically work between 7am and 3pm, it is best to hold a training that mirrors those hours. People typically have commitments outside of work that do not allow them to shuffle their schedules. It is important not to inconvenience people.

Consider the Appropriateness of Facilities. Selecting appropriate facilities is also an important aspect of designing an effective training. People learn best in comfortable environments.¹¹ Training organizers need to be sure there is enough space available to accommodate all of the participants. It is also important that the location of the training has the resources necessary to accommodate the presenters. Because adults learn best via multiple teaching strategies, many instructors are increasingly relying on audiovisual aids to enhance their presentations. It is important for training organizers to select facilities that allow instructors to utilize computers, LCDs, projectors, and/or DVD and stereo equipment. Failure to do so detracts from instructors' ability to tailor their presentation for multiple learning styles, and could limit the utility of the training and ultimately the evaluation.

Select Appropriate Faculty. It is important to establish the credibility of instructors. Selecting faculty with high levels of experience or scholarly expertise can go a long way in ensuring that trainees attend to the presenter's materials. Authority figures should be selected whenever possible. The best instructors are highly familiar with the information they present and are able to establish their credibility when challenged by trainees. Credible instructors are vital to motivating trainees to listen and learn, thereby increasing the likelihood that a training

⁸ Kirkpatrick, D.L. & Kirkpatrick, J.D. (2006); Knowles, M.S., Holton, E.F., & Swanson, R.A. (2005).

⁹ Knowles, M.S., Holton, E.F., & Swanson, R.A. (2005).

¹⁰ See for example Beaudin, B.P. and Williams, R.E. (1990); Brookfield, S. (1992); and Cross, K.P. (1981).

¹¹ Kirkpatrick, D.L. & Kirkpatrick, J.D. (2006).

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

evaluation will be effective. In addition to having command of the material, it is effective to select instructors who are familiar to the trainees or with the trainees' environment and organizational structure and issues. The more a training instructor appears in control, the more credibility he or she will have with the audience.

SECTION 1.4

Preparing Adults to Learn

Early learning theorists challenged the long-held assumptions that people do not learn much after childhood and re-directed discussion toward how adults learn.¹² Research on preferences and styles of learning also advanced our knowledge of adult learning.¹³ Research focusing on adult learning expanded greatly in the 1960s and 70s with the infusion of insights from psychology and further exploration of the differences between adult and child cognition. Key assumptions about adult learning emerged from this research; immediacy, self-direction, experience and motivation. Knowles (1980; 1984; 1998), Vella (1995) and others, popularized these principles resulting in broad dissemination and adoption into practice.

Understanding and utilizing the principles of adult learning is essential to program planning and design, regardless of the training topic area, faculty selected, etc. Attention to the principles of adult learning not only suggests areas for evaluation focus, but will also impact the overall effectiveness of a training program.

Table 1-6	
Principles of Adult Learning: Cognitive Psychology	Principles of Adult Learning: Education
Immediacy – adults use new information and skills to solve immediate challenges	Need to know – adults value training when needed for practical reasons
Self-direction – adults identify their learning needs and pace themselves	Self-concept – adults are independent and self-directed
Experience – adults have a reservoir of experience and insight to bring to the learning environment	Life experience – adults have accomplishments and knowledge to apply to learning
Motivation – adults are internally motivated to learn rather than dependent upon external motivation	Task-centered and practical – adults value learning that meets job-related needs
	Internally motivated – adults are motivated by internal rewards and not grades or parental approval

¹² See for example Thorndike, E.L. (1929).

¹³ See for example Jung, C. (1971).

Principles of Adult Learning¹⁴

Adults learn differently than children and adolescents. Unlike children who believe that their teachers are omnipotent role models, adults need to be persuaded and motivated to learn from an instructor. In order to design and implement effective trainings for adults, trainers need to abandon the traditional pedagogy model (i.e., lecturing) used to educate children and young adults in institutions of learning (i.e. grammar school through college) and embrace the principles of adult learning, termed the andragogical approach.¹⁵ According to this approach, there are six primary principles that impact adult learning, and consequently the overall effectiveness of a training program.

Andragogy – “any intentional and professionally guided activity that aims at a change in adult persons”

Knowles et al., (2005)

Building on Andragogical Assumptions, the SIX PRINCIPLES that Impact Adult Learning, in Training Program Planning are:¹⁶

1. **The human “need to know.”¹⁷** People need to know why it is important to learn the information that is being presented to them. Adults inherently ask themselves “why is this important”? Answering this question effectively is essential to ensure that trainees will be motivated to listen and comprehend the information being presented to them. Adults have many responsibilities and can be impatient and resistant to learning if their time is wasted. Establishing the utility and necessity of any training is an essential first step in determining the program’s effectiveness.

TIPS

- Begin and end your session on time.
 - Understand who is in the audience and why they are participating (e.g., ensure participants have an opportunity to introduce themselves).
 - Use a needs assessment process to learn what questions your audience has about the subject – Don’t cover material they already know unless there is a good reason for it.
 - Recognize that your subject is only one of many that participants may be interested in learning more about.
2. **The need to overcome the adult ‘self-concept’.¹⁸** Unlike children and adolescents whose perceptions of themselves are in flux, adults have a stable conceptualization of who they are. Adults perceive themselves as competent beings who are capable of making their own decisions and who are responsible for their own behavior. Adults are decision-makers, self-directed learners, and bring a great deal of experience and

¹⁴ Knowles, M.S., Holton, E.E., and Swanson, R.A. (2005).

¹⁵ Knowles et al., (2005).

¹⁶ Knowles et al., *supra* note 14

¹⁷ *Supra*, note 2.

¹⁸ *Ibid.*

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

knowledge to any learning situation. This leads adults to develop a deep psychological need to be treated as autonomous individuals. As a result, attempts to introduce new information or to alter behaviors may be met with resistance. Therefore, an effective training must be carefully constructed to ensure that participating trainees feel as though they have some influence in the learning process.

TIPS

- Through surveys, interviews, focus groups or a needs assessment process, provide opportunities for trainees to provide feedback about their training needs.
- Through surveys, interviews, focus groups or a needs assessment process, provide opportunities for trainees to provide feedback about their training needs.
- Through surveys, interviews, focus groups or a needs assessment process, provide opportunities for trainees to offer input about training venues, timeframes, mode of delivery, etc.
- Show respect for participants' experience by asking them to share ideas, opinions, and knowledge. Verbally recognize that they may be a good resource for reaching your teaching goals.
- If you already know the participants, you may recruit particular individuals who can provide you with helpful input before, during, or after the delivery of your curriculum.
- Be the "guide on the side" rather than the "sage on the stage."

3. The readiness to learn.¹⁹ Adults will automatically question the credibility of their instructors. Unlike children (and to a lesser extent adolescents), who are trained to blindly obey the teacher, instructors of adult trainings must establish their credentials. It is not uncommon for trainees to openly question instructors or facilitators during trainings, a process termed 'immediate feedback.' Establishing the credentials of facilitators and lecturers is vital in preparing adults to learn.

TIPS

- It is imperative that instructors establish their credentials at the beginning of the training (e.g., through provision of biographical statements in training materials and through introductory comments which highlight their education, background, and experience).
- Pair expert faculty on a particular practice area with expert practitioners (e.g., team psychological or medical experts with judges who can make the links from content to practical application in the court context).

4. Experience of the audience and group impact.²⁰ It is important to recognize that people have varying levels of exposure to and experience on any given topic. Trainees vary in their level of education, the time spent on the job, and their commitment to their occupation (i.e., full time vs. part-time employees, intrinsically interested vs. other interests). Knowing the trainees' level of experience helps organizers determine the

¹⁹ Knowles et al., *supra* note 14.

²⁰ Arnold, W.E. & McClure, L. (1996); Knowles et al., (2005).

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

substance of the training curriculum. People that have little experience regarding the information provided in a training program will need to be introduced to the topic. In this scenario, the substance of the training will be more informative. People with more experience are likely to be familiar with the topic, allowing for the information provided in the training to be more action-oriented. Understanding the level of experience trainees have on the topic can direct organizers as to what goals are appropriate for a training program, and how to design and interpret evaluation findings.

In addition, adults are also often concerned that participating in a group will make them look weak, either professionally or personally. In order to provide the most comfortable learning environment, and ensure that adult learners are receptive to the curriculum, training organizers should provide opportunities and allow time for people to establish themselves in the group before implementing group activities as part of the learning.

TIPS

- Use a training needs assessment process to gain knowledge of your audience's interests, motivation, skills, and training needs. Use findings from the needs assessment to target workshops, content, and exercises to the mixed needs of your trainees. See Chapter 2.5 in this Section for more on training needs assessments.
- When your audience has mixed levels of experience and knowledge, acknowledging those differences to the larger group will justify why there is content aimed at an introductory level as well as content aimed at intermediate or advanced levels.
- When your audience has mixed levels of experience and knowledge, ask those individuals with greater knowledge to offer their experiences, lessons learned, etc. Use those individuals as coaches or mentors for other trainees.
- Design training workshops, educational exercises, and discussion sessions that help people feel safe enough to ask questions and confident that they will be respected.
- Don't ask people to take risks too early in a workshop or course (for example, engaging in a role play exercise) unless they already know each other well.

5. **The need to motivate people.**²¹ For adults to change their behavior, it is imperative that trainees feel as though there is a direct benefit from the training.²² Adults are motivated by information or tasks that they find meaningful. Common motivators include monetary and non-monetary rewards, security, power, prestige, happiness, harmony, meeting organizational requirements, meeting organizational goals, or easing workloads. Exactly which motivator works best in any given training requires an understanding of the needs and goals of the audience. It is incumbent upon the training organizer and instructors to

²¹ Knowles et al., *supra* note 14.

²² *Ibid.*

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

persuade trainees that the information presented directly benefits them and their ability to do their job effectively.

TIPS

- Use a training needs assessment process to learn more about your audience's motivation for attending the training. Use the needs assessment to become aware of what people want (and need) to learn, how much they already know, and the kinds of generative themes that might affect their attention.
 - Generative themes are concerns and issues that are most important in a person's life.
 - Generative themes may enhance or challenge a person's ability to learn.
 - Generative themes could include the desire to do a better job, the need for change, new possibilities for achieving positive outcomes, ways to overcome a problem or challenge, or opportunities to share experiences.
- Offer continuing education credits for training attendance and completion of the training.
- Offer individuals who will complete the training an opportunity to teach in future training programs.

6. **The recognition that different people have different learning styles.**²³ Peoples' ability to retain information is enhanced by using combinations of various learning styles. People can learn information by reading, listening, observing, or participating. Which method works best varies by the person, but there is some evidence as to which works best *in general*. A classic study by the US Department of Health, Education, and Welfare (1974) found that people typically remember only 10% of what they read and 20% of what they hear, 50% of what they see and hear, and 90% of what they say and do. Generally, the most effective of the four learning styles is participation (see figure 7); however, it is important to incorporate each learning style for maximum effectiveness, as it allows people to focus on the material that makes the most sense for them. Therefore, it is best to employ a combination of teaching methods that tap into different learning styles.

²³ Arnold, W.E. & McClure, L. (1996); Kirkpatrick, D.L. & Kirkpatrick, J.D. (2006); Knowles et al. (2005).

Categories of Learning Experiences or Styles²⁴

- **Doer** – Likes to be actively involved in the learning process, wants to know how he or she will apply learning in the real world, likes information presented clearly and concisely.
- **Feeler** – People-oriented, expressive, focuses on feelings and emotions, thrives in open, unstructured learning environment.
- **Thinker** – Relies on logic and reason, likes to share ideas and concepts, analyzes and evaluates, enjoys independent work.
- **Observer** – Likes to watch and listen, tends to be reserved, will take his or her time before participating, and thrives on learning through discovery.

TIP

- Use a checklist process for planning – have you hit each learning style in your training plan, associated agenda, training materials, and breakout exercises?

Training Methods and Learning Styles

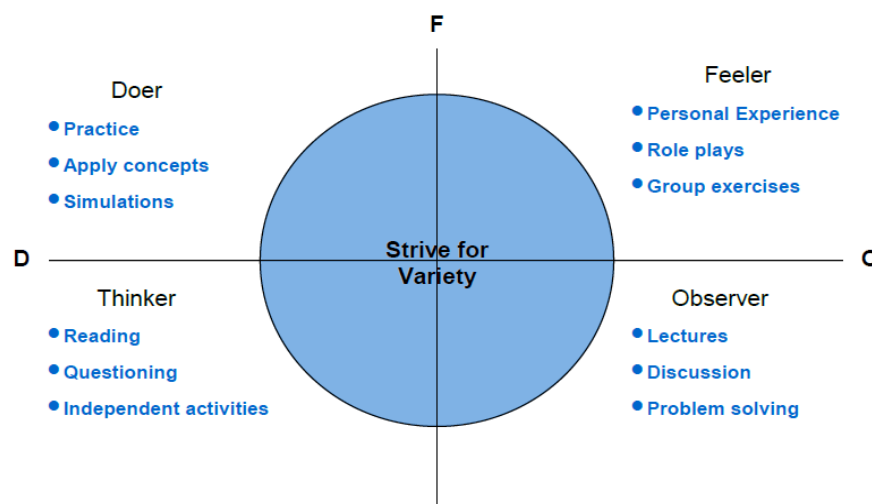


Figure 7

²⁴ Lawson A.E. & Johnson, M. (2002); Kolb, D.A. (1984) – in Kolb's original work the learning styles were: Converger (favoring abstract conceptualization and active experimentation), Diverger (favoring concrete experience and reflective observation), Assimilator (favoring abstract conceptualization and reflective observation), and Accommodator (favoring concrete experience and reflective observation).

Matching Learning Needs and Styles to Training Methods

Table 1-7	
Adults learn best when	TIPS for matching adult learning needs with appropriate methods
They feel valued and respected for the experiences and perspectives they bring to the training situation.	Elicit participants' experiences and perspectives through a variety of stimulating activities. Ask for participants to share their own experience and lessons learned with the larger class.
The learning experience is active rather than passive.	Actively engage participants in their learning experience through discussion and a variety of activities.
The learning experience actually fills their immediate needs.	Identify participants' needs; develop training concepts and learning objectives to these identified needs.
They accept responsibility for their own learning.	Make sure that training content and skills are directly relevant to participants' experiences so that they will want to learn.
Their learning is self-directed and meaningful to them.	Involve participants in deciding on the content and skills that will be covered during the training.
Their learning experience addresses ideas, feelings, and actions.	Use multiple training methods that address knowledge, attitudes, and skills.
New material relates to what participants already know.	Use training methods that enable participants to establish this relationship and integrate new material.
The learning environment is conducive to learning.	Ensure that the physical and social environment (training space) is safe and comfortable.
Learning is applied immediately.	Provide opportunities for participants to apply the new information and skills they have learned.
Learning is reinforced.	Use training methods that allow participants to practice new skills and receive prompt, reinforcing feedback.
Learning occurs in small groups.	Encourage participants to work in small groups to explore feelings, attitudes, and skills with other learners.
The trainer values participants' contributions as both learners and teachers.	Encourage participants to share their expertise and experiences with others in the training

Table 1-8	
TIPS for Matching Learning Styles with Instructional Method	
Learning Styles	Consider using:
Learn best with abstract concepts and lectures	Case studies and discussions about theories and research
Learn best while observing others	Demonstrations, simulations, and filmed reenactments
Learn best from exercises	Role playing and other experiential activities
Learn best through visual means	Videos, images, and slides

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

In summary, the principles of adult learning should guide the process you use when training professionals. The following key factors will help you conduct successful training programs:²⁵

1. Because learning is enhanced by challenge and inhibited by threat, establish an environment where participants feel safe and supported; where individual needs and uniqueness are honored, and abilities and life achievements are acknowledged and respected.
2. Treat participants as peers, who are accepted and respected as intelligent experienced adults whose opinions are listened to, honored, and appreciated.
3. Deliver content that has relevance, so participants can relate what they learn to past experience. The brain automatically searches for meaning, patterns, and relationships based on prior knowledge and experience. Learners cannot separate the learning of a skill from the meaning that skill has for them.
4. Learners will perceive the content within their own global view. Therefore, elaborate on your intended context, including history, purpose, methods, and intended results.
5. Deliver content that has immediacy; people learn best what they can apply right away.
6. Facilitate self-directed learning, where participants make action plans and take responsibility for their own on-going, professional development.
7. Provide opportunities for participants to give feedback and input to the learning process, and to give and receive feedback and input from trainers and other participants.
8. Guide learning processes that foster team work and provide opportunities for team members to:
 - Develop team norms and guidelines for working together
 - Share perspectives, knowledge, insight, and experience
 - Tell personal stories, creating common ground, and connection
 - Develop materials and carry out tasks
 - Make action plans and decisions
 - Have fun together
9. Learning involves both focused attention and peripheral perception. Trainers should pay attention to all facets of the educational environment.
10. Provide learning processes that require active involvement. Provide opportunities for real:
 - Problem Solving
 - Practice of judgment skills
 - Reflection and inquiry
 - Intuitive reasoning
 - Interactive questioning
 - Learning and practicing critical thinking skills
 - Meaning exploration
 - Exploration of values and feelings
11. Foster intellectual freedom and encourage experimentation and creativity.

Consideration of principles of adult learning and learning styles is necessary to overcome the obstacles that are associated with teaching adults. These factors need to be considered and controlled for in order for a

²⁵ See Renate, M. & Caine, G. (1994).

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

training to be effective. These principles should be applied independent of the topic being discussed, or the context in which a training occurs. Not doing so will leave organizers with an audience unwilling to listen and learn. The effectiveness of the adult learning methods applied should also be evaluated (e.g., how satisfied were participants with the methods used? Did the methods employed result in gains in knowledge or skill? Did the methods employed facilitate behavior or attitude change?).

Adult Learners:

- Have a good deal of first-hand experience that they wish to use and share in class.
- Expect to be treated with respect due their maturity and individualism in the learning situation.
- Usually have specific and immediate learning goals and expect structure and clear outcomes for the learning program. They want to know how to apply learning to their personal or professional lives.
- Have a desire to be active participants in the learning process. Effective learning situations are interactive and tend to be centered on problem solving.
- Are frequently anxious about their learning abilities and the appearance of competence in the classroom, but are anxious for educational success.
- Have a strong need for periodic feedback, encouragement, and learning in an atmosphere where there is a high degree of safety, mutual commitment, and choice.
- Are critical of unprepared teachers, poorly articulated programs, and individuals or processes which interfere with their learning.
- Expect to have their physical needs met (adequate furniture, appropriate breaks, etc.).
- Need a good balance between tight, well-paced, content-oriented presentations and the time needed for learning integration.

SECTION 1.5

Planning Training with an Evaluation Focus

Using Inside or Outside Evaluators

When planning for a training evaluation, you should consider whether the evaluation should, or can be, conducted internally (e.g., by training or organizational staff), or if an outside evaluator should be sought. Level four (results or outcome-based) evaluations, for example, may require expertise in research design, methods, and analysis that is lacking in internal evaluation staff. While it may not be completely necessary to obtain an outside evaluator in these circumstances, training organizers may want to outreach to a local university, college, or research community for technical assistance in developing and implementing an evaluation plan, including data collection and analysis strategies.

Whether evaluators are insiders or outsiders, it may be helpful to have in place an advisory group or committee to provide input into the evaluation plan and its implementation. Advisory group members should include individuals with intimate knowledge of the dependency court system, individuals responsible for the design and implementation of training programs, and individuals with research and evaluation experience.

In making a decision about using an inside or outside training evaluator, training organizers should consider the following factors:²⁶

- **Administrative Confidence** – Stakeholders and other recipients of training evaluation reports may not have confidence in the evaluation skills of internal staff. Competence is a big factor in ensuring confidence and training organizers should ensure any inside evaluators have the necessary skills and knowledge to design and implement the level of training evaluation required. In addition, it is often important to ensure public confidence in evaluation findings by engaging researchers who have no stake in the training being studied.
- **Objectivity** – Objectivity requires evaluators to be free from bias in their interpretation of findings. Outside evaluators who have no stake in the outcome of the evaluation have an advantage over inside evaluators with respect to objectivity (or at least the public perception of objectivity). While no evaluator (inside or out) is totally objective (we all have beliefs and values which inevitably color our methodological choices and interpretations), training organizers should seek conditions that minimize biases for or against the training program.

²⁶ Lipsey, M.W. (1985); Weiss, C.H. (1998).

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

- **Understanding of the Training Program** – Knowledge of the training program, its goals and learning objectives, is critical for an evaluator. In-house evaluators who have been involved in the development of the training or who are in a position to understand the real issues that the training is designed to address have an advantage over outside evaluators in terms of their understanding. If outside evaluators are used, training organizers should ensure that they are educated about training program processes and make the effort to access sources of information about the training.
- **Potential for Use** – In-house evaluators, who report results and make recommendations on the basis of results, have many opportunities to bring training evaluation findings to the attention of stakeholders (both line-level stakeholders and those in leadership positions). However, sometimes it is outside evaluators who have the necessary credentials to induce stakeholders to pay attention to the evaluation findings.
- **Autonomy** – Outside evaluators, who are not entrenched in the “usual” or typical way of doing things in the organization, may be able to suggest and implement more creative approaches to conducting a training evaluation. With respect to the analysis of findings, outside evaluators may be more oriented to fundamental rethinking of program parameters – helping training organizers see results in a broader context.
- **Balance** – Training evaluation can be performed by internal staff or by outside evaluators. Each has its advantages and all of the above factors should be balanced against each other when making a decision about using an inside or outside evaluator. Inside evaluators understand the organization and issues the training is designed to address. They also understand organizational interests and needs and have opportunities for putting evaluation results to practical use. Outside evaluators have greater autonomy, often greater prestige, and often have a wider range of skills. Cost and skill required to design and implement a particular evaluation approach are critical deciding factors.

While using skilled evaluators is an important prerequisite for quality evaluation, there is room for diversity in the backgrounds and knowledge of evaluators. Of course, much depends on the sophistication of the evaluation approach under consideration, but in general, anyone tasked with evaluating a training program should be knowledgeable about the organizational context and issues the training is hoping to impact. Evaluators should be aware of the many evaluation strategies available and how they may be matched most effectively to evaluation questions. Evaluators also need a variety of skills to be effective – they should be good analysts, good listeners, and possess excellent oral and written communication skills. Ideally, evaluators should be involved at all stages of the training program including program planning and design stages.

Evaluator Ethics

As the field of evaluation became increasingly professionalized, many associations published standards that could guide them in their evaluation work and ensure good evaluation practice. For example, the *Guiding Principles for Evaluators* promulgated by the American Evaluation Association (AEA) were developed in 1994 as guidelines for sound, ethical evaluation practice.

When working with outside evaluators, dependency court training organizers should ensure that those evaluators are not only familiar with social science research methods and the dependency court context, but also with the ethical standards of evaluation practice.

These *Guiding Principles* have been broadly vetted with the AEA membership and are reviewed and revised at regular intervals in order to ensure that they remain current with the field. The *Guiding Principles* set out five general areas for evaluators to follow in practice (each of these five principles is elaborated on in detail in the *Guiding Principles* which are available at www.eval.org):

- A. Systematic Inquiry - evaluators conduct systematic, data-based inquiries about whatever is being evaluated.
- B. Competence - evaluators provide competent performance to stakeholders.
- C. Integrity/Honesty - evaluators ensure the honesty and integrity of the entire evaluation process.
- D. Respect for People - evaluators must respect the security, dignity, and self-worth of the respondents, program participants, clients, and other stakeholders with whom they interact.
- E. Responsibilities for General and Public Welfare - evaluators articulate and take into account the diversity and values that may be related to general and public welfare.

In addition to the American Evaluation Association ethical standards, the American Psychological Association (www.apa.org) and the American Educational Research Association (www.aera.net) provide ethical guidance to the evaluation community.

The Training Needs Assessment

An important tool in training program design and evaluation is the training needs assessment. Conducting a training needs assessment will help you to determine the needs of your training audience and how best to apply the different adult learning principles to your audience and training content. It will help determine your audience's motivations for the training and whether there are any differences in their readiness to learn. Training needs assessments, if structured to obtain information about gaps in knowledge, pre-existing behavioral patterns and attitudes, can also provide important baseline data that can then be compared to questions about knowledge, behavior, and attitudes post- training.

Questions to be Answered by a Needs Assessment Process:

- What do the participants need to know and do as a result of this training?
- What do training organizers need to know about the course participants and their jobs/responsibilities?

Needs Assessment: Definitions and Context

A Need – A “need” refers to the gap between what is and what could or should be within a particular context, leading to strategies aimed at eliminating the gap between what is and should or could be.

A Needs Assessment – A program-based needs assessment is a systematic inquiry that helps to identify program priorities and helps make program implementation decisions. A program-based needs assessment helps to allocate finite resources in a manner consistent with identified program goals and objectives.

A Needs Assessment Process Includes:

- Identifying and analyzing expressed and unexpressed needs
 - Expressed needs – individuals recognize the need for increased knowledge and skill in specific areas and express a desire to pursue education. This results in the design of training programs that address participants' demands or requests for this knowledge.
 - Unexpressed needs – individuals lack awareness of their need for competency and education in specific areas and do not actively pursue training in these areas. This results in the need to implement an awareness campaign so that individuals will recognize the need.
- A plan to develop strategies that address such needs.

Who? ... Who needs the training program? Judges generally or judges who meet some specific requirements (new judges, specialized judges, multi-disciplinary partners)?

What? ... What are the specific training goals? What areas of practice, policy, or research need to be included in the training program? What are the best teaching strategies given the topic and target trainees? What faculty would best serve the training program goals? What material is necessary? What does the training cost? Overall, what are we trying to achieve through the training program?

When? ... When should the training program occur? How often should the training program occur? How should time frames – from specific training components to the overall training agenda – be determined?

Where? ... Where will the training program be held?

Why? ... Why are we conducting training programs in identified areas? Why were the specific training programs area identified?

A training needs assessment measures what is currently in place and what is needed, now and in the future. A needs assessment provides you with an understanding of current shortcomings and aids in the decision-making process by defining all elements, issues, facts, and practices that need to be taken into account when designing your training. Trainings have failed in the past, and will continue to fail because training designers did not understand the needs of the consumers of the training.

A training needs assessment can also build understanding in the consumer of the training about the learning activities and their purpose. During the needs assessment you are bringing potential learners into learning design activities and making them part of the process. Besides introducing learners to the training activity, learners will also accept and benefit from a training program that they themselves helped to define.

A formal needs assessment may not always be needed for evaluation (although it is ideal). If time and resources do not permit a formal needs assessment process, training organizers can build informal needs assessment into the evaluation strategy. Before conducting the program, for example, sample a small segment of your target training population and collect baseline data about behavior, practices, and knowledge whenever possible – look at existing records for hard data. A focus group is also a quick way to gather this data. You can administer a brief questionnaire at the beginning of the program itself (at the beginning of the training). Be sure to sample the same group you targeted in the initial assessment process for your post- training evaluation study.

Key Tasks of the Needs Assessment Process

Determine the target population

- Identify what type of individual or group(s) the training program is designed for (one group? Multiple groups? Multi-disciplinary?)

Understand the participants' characteristics

- Experience
- Cultural background
- Education
- Location
- Mindset/Motivation
- Challenges (location, job demands, resources, etc.)

Determine the participants' needs

- Conduct a literature review: analyze existing performance data, mission and goal statements, evaluation reports, or other documents for existing problems and challenges
- Draw from your past experience with similar groups
- Gather information from informal discussions among appropriate participant groups
- Conduct surveys
- Conduct focus groups composed of stakeholders
- Work with an advisory panel
- Observe participants – watch stakeholder performance in the field
- Interview potential training participants, managers, and subject matter experts
- Learn about critical incidents
- Determine what emerging data should be distributed

Key Factors for a Successful Needs Assessment

- Keep in mind the value and necessity of broad-based participation by stakeholders
- Choose appropriate means of gathering information about critical issues
- Recognize core values in the group whose needs are being assessed
- Needs assessment is a participatory process; it is not something that is “done to” people
- Needs assessment cannot ignore political factors – the priorities that are derived may be counter to accepted, standard or entrenched ideas in the system
- Data-gathering methods by themselves are not a needs assessment – the needs assessment is a total decision-making process in which the data are but one component

Some Creative Training Needs Assessment Strategies

Critical Incident Technique – Your needs assessment survey or questionnaire can use a “critical incident technique (CIT),” either in part or as the whole survey.²⁷ CITs are direct observations or self-report of specific

²⁷ Flanagan, J.C. (1954). The Critical Incident Technique. *Psychological Bulletin*, Vol. 51 (4), p. 327-359; Fivars, G. (Ed) (1980) *Critical Incident Technique*, American Institutes for Research.

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

behaviors that relate to performance in a given situation. For example, ask respondents to recall specific events or recent conditions observed that illustrate the dependency court system is doing an unsatisfactory job or something about the system that needs improving. Then, ask respondents to recall critical incidents that show that the system is doing a satisfactory job. Ideally, you would survey a variety of respondents to elicit critical incidents from a diverse group of stakeholders. The needs assessor sorts the positive and negative incidents according to program areas or issues. This technique may elicit goals that may be difficult to obtain using standard question/ answer formats (e.g., problems in initiative, responsibility, resourcefulness, motivation, empowerment, leadership, etc.).

Stakeholder Forum – Conducting a forum is a well-established procedure for needs assessment. In general, it's analogous to a town hall meeting in which a community of stakeholders is called together to discuss issues. In the needs assessment context, the forum is used to gather stakeholder concerns or perceptions of need areas, opinions about the quality or delivery of system responses and services, information on causes of needs, and exploration of stakeholder values. One purpose of the forum might be a review of goals that are being considered as performance targets or practice standards. For example, participants can be provided with a copy of goals in advance of the meeting, as well as a few questions guiding the examination of the goals.

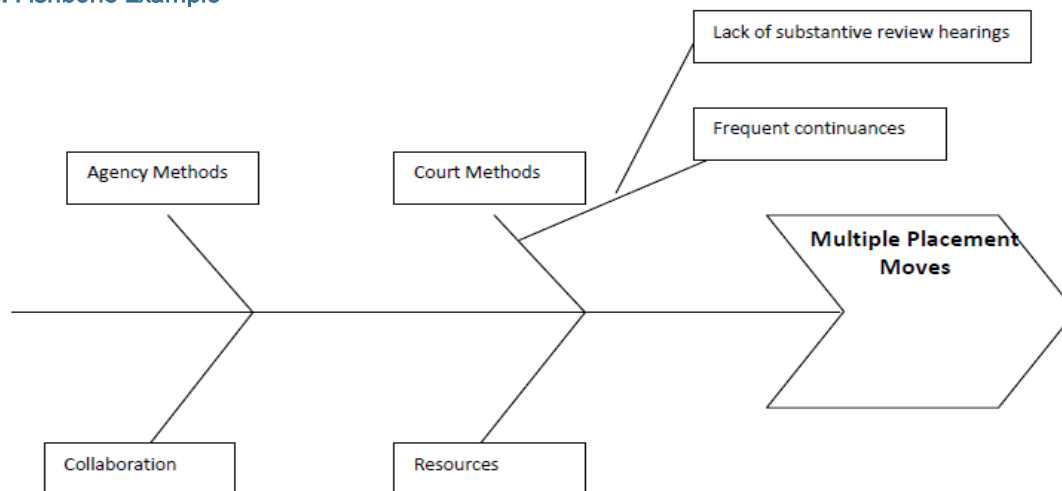
Choosing participants for the forum should be based on whether you want a cross-section of the stakeholder community; if there are specific groups that you need to include in order to discuss specific issues; and what you hope to gain from their involvement – would key informants from particular groups suffice? As a general rule of thumb, diversity in sampling is often the best strategy to follow, especially when your goal is generating a range of opinions and perceptions about a need area.

Scenarios – Scenarios are pictures or snapshots of what some specific future event or system will look like. In training needs assessment, if we have a vision of a target state and compare it to the present situation, we may illuminate needs or discrepancies that need resolution through training. A group of key system informants is asked to develop a hypothetical word picture of the dependency court system in some specified time in the future (e.g., one year from now, five years from now, etc.), including desired goals, objectives, relationships, communication patterns, skill-sets, and outcomes. When using scenarios, it's important to select participants carefully because they are crafting a future vision of the system which will be used by the needs assessor to identify training needs. Consider using teams of leaders, key decision-makers or managers; or consider using a number of mixed or multidisciplinary stakeholder groups; or convene scenario-groups by role (e.g., a group of judges, a group of agency attorneys, etc.). Once your groups have defined the ideal system scenario, they are asked to identify the factors that are within the system currently that are likely to increase or decrease its chances of achieving the future scenario's desired goals and objectives. Needs assessors then analyze the results to identify training areas that will help achieve future scenarios.

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

Fish-boning – Fish-boning is a technique for identifying the causes of existing needs,²⁸ which in turn can generate a list of education and training priorities. This technique is used in small groups, such as a focus group, of usually 10-12 individuals who are familiar with the need or problem area. More than one group may be used, either at different times or concurrently but in different rooms or different sections of a large room.

Figure 8: Fishbone Example



A large fish diagram (figure 8 above) is provided to the group on a flip chart. The fishbone displays the general need or problem, and four or five “ribs,” selected by the needs assessor, to represent possible key factors related to causes of the problem. For example, a group may be asked to address the problem of multiple placement moves for children. This problem is written on the fish’s head, with the ribs identified with the labels of “court methods”; “child welfare agency methods”; “resources;” and “communication/ collaboration.” Each group is asked to write down as many examples of the causes of the problem or need as possible, noting them as radiating lines or twigs that extend from the ribs of the fish skeleton. For example, a group may identify “frequent court continuances,” and “lack of substantive review hearings,” as two “court methods” that are contributing to the problem of multiple placement moves. An alternate method would be to ask the groups to draw in the ribs and label them themselves (e.g., the group members identify the root causes of the problems without guidance from a facilitator). When using this approach, groups may identify a large number of “ribs” or possible causes of the problem. Facilitators would then have to engage the groups in prioritizing among those causes in order to narrow the focus of the needs assessment. Once groups have completed their fishbone diagrams they share them with the group facilitator and other groups. Needs assessors can use the information about perceived causes of problems to identify and prioritize training.

²⁸ Axon, D.A (2007). *Best Practices in Planning and Management Reporting, Second Edition: From Data to Decisions*. John C. Wiley and Sons.

Tips for Conducting Useful Needs Assessments

- Invite stakeholders to share their thoughts on what they believe training participants will likely need.
- Identify stakeholders who have already expressed an interest in the training and identify their learning needs through a short survey or interview.
- Use available needs assessment data already gathered (i.e., what do you already know about the target population?).
- Determine in general at what level the training will be taught (trainees are new to role, moderate experience, or very experienced).
- Ask presenters/trainers who have been involved in similar trainings what they believe are the most important skills that the target learners need to develop.
- Find other courses held in other jurisdictions on a similar topic and identify topics covered.
- Search topics in recent journals, technical assistance publications, etc., for timely and relevant content areas related to your target population's learning needs.
- Physically observe stakeholders in the field to gain knowledge of their practice.
- Interview stakeholders and ask them what kinds of training they need to help them become better and more effective in their roles.
- Try to gather information from system consumers (e.g., parents, foster parents, children and youth) that will help you identify what skills should be taught during the training.
- Help participants identify what they don't know and what they need to develop greater competency in. Accomplish this by asking them to complete a questionnaire that focuses on their present job tasks and what they need to learn to improve their knowledge, skills, and sometimes, confidence.
- Review previous training evaluations to determine what areas need to be emphasized and what topics require less time.
- Identify barriers and constraints to conducting the training; determine which ones you can manage and which ones you need assistance with.
- Based on what you are learning about your target population and their learning needs, try to approximate how much time is necessary to teach each identified topic.

SECTION 1.6

Designing Measureable Training Objectives

A critical part of planning trainings with an evaluation focus is the identification of measureable training or learning objectives. This process involves identifying learning goals and learning objectives. If you have completed a training needs assessment, learning goals, and their associated objectives, should be tied to the findings of your assessment. When identifying learning goals for a training event, you can ask yourself “What do you want participants to have learned by attending this training?” “What attitudes and behaviors do you want to see changed as a result of this training?” “What policies or procedures do you want to see implemented as a result of this training?” “What practices do you want to see improved as a result of this training?” And, ultimately, “What impacts do you want this training to have on participants, the systems they represent, and the children and families they serve?”

Each learning goal should have an associated learning objective. A learning objective statement should be a clearly worded sentence that expresses what participants will have learned and how they will demonstrate their learning (e.g., “By the end of this training a participant should be able to ...”). Writing clear learning objectives answers the question, “What will the learners be able to do when they finish the training program?” Of all the activities within the training design process, this is one of the more critical steps. For without well-constructed learning objectives, instructors don't know what is to be taught, learners don't know what they are supposed to learn, managers don't know what they are investing their training dollars in, and evaluators won't know what they are measuring. Learning objectives form the basis for what is to be learned, how well it is to be performed, and under what conditions it is to be performed.

A learning objective is a statement of what the learners will be expected to do once they have completed a specified course of instruction. It prescribes the conditions, behavior (action), and standard of task performance for the training setting.

Differences between Goals and Objectives

Goals describe a learning outcome in general; for example, “to improve child protection hearing practice.” Notice that this goal is so general that it provides no guidance about *how* it is to be achieved. On the other hand, an objective is a specific statement of instructional intent which attempts to

Goals are broad; objectives are narrow.
Goals are general intentions; objectives are precise.
Goals are intangible; objectives are tangible.
Goals are abstract; objectives are concrete.

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

change knowledge, skills, attitudes or behaviors, as a result of a learning experience, for example, "learners will understand how to apply the strategies for engaging parents covered in the training in their day-to-day practice."

In many situations the words "goals" and "objectives" are used interchangeably. Yet, in the context of goal setting, the difference between goals and objectives has an important practical meaning. A program goal is a statement that explains what you wish to accomplish. A program goal sets the fundamental, longer range direction – it expresses a long term aim you wish to achieve. A goal offers the "why" or explanation behind a program activity or task. Goal Example: "To improve court and child welfare agency collaboration." An objective breaks down goals into their smaller parts. Objectives are concrete attainments that can be achieved by following specific steps. Objectives may provide guidelines for how a goal can be accomplished. Objective Example: "To hold a multi-disciplinary stakeholder meeting once per month to discuss cross-system issues."

A Learning Objective is:

- A concrete and specific statement that explains what will be accomplished in order to fulfill organizational goals.
- A measureable level of achievement.
- A purpose statement – providing clear expectations, guiding content and helping to organize tasks
- A purpose statement that provides a basis for evaluation

SMARTER Training Objectives

SMART or **SMARTER** is a mnemonic (a memory or learning aid) used in project management and performance measurement to develop project objectives.²⁹ It is a helpful way to ensure that objectives are measureable. "SMARTER," for example, is an acronym for Specific, Measurable, Acceptable, Realistic, Time-Bound, Extending, and Rewarding. When developing your training objectives you should ensure that they are:

Specific – Specific objectives are well-defined and clear to anyone that has a basic knowledge of the program or project. For example, it's difficult to know what someone should be doing if they are to pursue the goal of "work harder." It's easier to recognize "meet with clients more frequently."

Measurable – Objectives should be obtainable and specify how far away that attainment or completion is – you should be able to know when your objective has been achieved. It's difficult to know what the scope of "meet with clients more frequently" really is, for example. But it's easier to determine if the objective has been achieved if it's written as "meet with clients a minimum of once per week."

Acceptable – All relevant training stakeholders should understand and accept the training's learning objectives. If a trainee is to take responsibility for the pursuit of an objective, for example, the goal should be

²⁹ Doran, G.T. (1981); Pohl, M. (2000).

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

acceptable to that trainee. A trainee is not likely to follow the directions of someone telling him/her to meet with clients a minimum of once per week if he or she does not believe that the objective is worthwhile. However, if your trainees understand why the objective is worthwhile, and why it is relevant to their role and practice, they are much more likely to accept pursuit of the objective.

Realistic – Objectives should be within the availability of resources, knowledge and time. Even if trainees do accept responsibility to pursue an objective that is specific and measurable, the objective may be written in such a way that it would be impossible to attain given current resources. The goal of “meeting with clients a minimum of once per week for a minimum of four hours each visit” may not be realistic. In this example, a more potentially achievable minimum time allotment for client meetings needs to be considered for the objective to be realistic.

Time-bound – Specifying a time frame for the objective’s attainment not only helps you recognize when the objective is achieved but also aids in evaluation. It means more to others (particularly if they are planning to help or guide individuals to reach objectives) if there is specificity about when an objective is considered complete. For example, “Within 30 days of completing the training, trainees will be meeting with clients a minimum of once per week.”

Extending – Ideally, some of the objectives you identify for your training program should resonate emotionally with trainees – trainees should care about, or feel something for, the objective. If you have included some objectives that individuals care about then they are more likely to commit to accomplishing those objectives. For example, trainees might be more committed to a strategy of meeting with clients a minimum of once per week if they understand how those meetings can benefit their clients (“In order to empower clients to active and early engagement in case plans, trainees will meet with clients a minimum of once per week”).

Rewarding – Ideally, some of the objectives you identify for your training program should also provide an example of the successes or rewards that can be attained if the objectives are accomplished. A trainee may be more inclined to meet with clients a minimum of once per week if he or she understand, for example, how that effort will make their job easier (“improve frequency and quality of communication”) and improve outcomes (“increase case plan compliance,” “timely access to appropriate services,” and “reduced time to achieve permanency”).

The Learning Domains and Developing Learning Objectives

Benjamin S. Bloom³⁰ developed a system of categorizing learning behavior in order to assist in the design and assessment of educational learning. In this taxonomy of educational objectives, the learner should benefit from development of knowledge or intellect (the Cognitive Domain); development of attitudes and beliefs (the Affective Domain); and the ability to put that learning into action (the Psychomotor Domain).

³⁰ Bloom, B.S. (1956); Pohl, M. (2000).

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

1. **Cognitive domain** (intellectual capability – knowledge, or 'thinking')
2. **Affective domain** (feelings, emotions– attitudes, opinions, 'feeling')
3. **Psychomotor domain** (skills – 'doing')

This taxonomy can be useful when developing learning objectives as you can examine whether your training's learning objectives tap each of the relevant domains. It can also be helpful in selecting the appropriate verbs to include in learning objective statements. See the Section Two Tools and Resources for guidance on selecting the right verbs for learning objective statements.

The following questions can help guide you as you develop your learning objectives:

1. Does your statement identify what learning domain it is in? For instance, is the focus on teaching information, improving a skill, changing an attitude or changing a behavior?
2. Is it realistic for your proposed audience given available resources, and the length of the training?
3. Is it focused on one specific job-related task or knowledge area?
4. Are there timelines that need to be considered and are they reflected in your learning objectives? Do you have to accomplish certain areas of knowledge or skill by a specific time, and if yes, are those times included in your learning objective statements?
5. Is the learning objective worded as a behavioral statement – something trainees actually have to do, not just understand or believe?
6. Is the statement specific and precise about what trainees are supposed to be able to do?
7. Is the statement written in a measureable way – can you measure it in some way to determine if learners have successfully demonstrated that outcome?

Be sure to determine your training goals and learning objectives yourself. Do not simply adopt goals from another training program. It can be very tempting to get a copy of a formal training program's goals and make them your own. While those goals may be similar or complementary, be sure to develop your goals based on knowledge of your audience and the specific content and scope of your training program. Learning goals should be established based on needed areas of knowledge and skills. These needs are established by referencing relevant strategic goals, objective data, competencies lists, knowledge about best practices, job descriptions, job analysis, tasks analysis, etc. If you want improved "leadership" or strengthened "collaboration" as goals, know what behaviors will depict "improved leadership" or "strengthened collaboration" and then identify those behaviors in your training plan.

Summary Rules for Specifying Learning Objectives

1. **Use strong verbs.** Use action-oriented verbs that describe an observable or measurable result that will occur. For example, "to increase the use of a judicial bench card in child abuse and neglect hearing practice," is an action-oriented statement involving a behavior which can be observed. In contrast, "to

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

promote greater use of a judicial bench card” is a weaker and less specific statement. The term “promote” is open to many interpretations.

2. **State only a single aim or purpose.** Most training programs will have multiple objectives, but within each objective only a single purpose should be specified. An objective that states two or more purposes or desired outcomes will require different implementation and assessment strategies, making achievement of the objective difficult to measure. For example, the statement “to begin a multi-disciplinary pre-hearing staffing program and provide more informative reports to the court” creates difficulties. This objective contains two aims – to begin a pre-hearing staffing program and to provide more informative reports to the court. It’s a “double-barreled” statement – if one objective is met (initiation of the program) but not the other (better reports to the court), to what extent has the overall objective been met?
3. **Specify a single end-product or result.** It is useful to have learning objectives with both a single aim and a single end-product or result. For example, the statement “to establish collaborative court-agency teams” indicates the aim but not the desired end product or result. Establish court-agency teams to what end or purpose? Including an end-result (“to establish collaborative court-agency teams who will produce a shared action plan for systems’ improvement”) provides more evidence for determining if the objective has been met.
4. **Specify the time of expected achievement of the objective.** It is useful to specify the time that the objective might be achieved. For example, to “establish collaborative court-agency teams as soon as possible” is not very useful because of the vagueness of “as soon as possible.” It is far more useful to specify a target date, or in cases where some uncertainty exists about some specific date, a range of target dates (e.g., “to establish collaborative court-agency teams between March 1 and March 30, 2011”).

SECTION 1.7

Using Logic Models in Training Evaluation

The Benefits of Logic Modeling for Evaluation

A logic modeling process helps you to articulate training goals, tie those goals to specific training activities, tie those activities to expected products or outputs, and tie those activities and outputs to desired outcomes. Having a training logic model in place helps you to ensure that training and training evaluation activities are specifically related to your theory of change – that training evaluation is logically tied to training goals and learning objectives and will provide you with information about whether those goals and objectives were achieved or not.

Logic Models help to illuminate a training program's theory of change by diagramming the “chain of events” - the training's expected events and expected “causal” linkages. Making explicit the theory underlying a program has several benefits for evaluators and for training organizers.³¹ A clear statement of theory illuminates the assumptions about behavior that underlie the intervention – in this case, the assumptions about behavior that underlie the training. Having a training program theory encourages evaluators and training designers to examine the empirical support for their assumptions about training impacts. Articulating the underlying theory clarifies what measures are needed to analyze the training program delivery and what outcomes to look for at what points in time, so that information about outcomes will be linked to data about the actions that produced the outcomes. Articulating the training program's theory of change also contributes to developing an understanding of what is needed for designing more effective future training programs in an iterative cycle of development, testing, and re-design.

Why Create a Logic Model?

Logic model development offers the following benefits:

- Clearly identifies program goals, objectives, activities, and desired results
- Clarifies assumptions and relationships between program efforts and expected outcomes
- Communicates key elements of the program
- Helps specify what to measure in an evaluation
- Guides assessment of underlying project assumptions and promotes self-correction

³¹ Chen (1989; 1990); Lipsey (1985).

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

Logic models have many applications for training programs. For example, during the training program design and planning stage logic models can be used to: clarify program strategy; find “gaps” in the theory or logic of programs; build a shared understanding about the program and how components work together; identify outcomes; and to establish timelines. During training program implementation logic models can be used to: develop a program management plan; emphasize connections between actions and results; make program adjustments; and to provide an inventory of assets and what is needed for program operation. During program evaluation and strategic reporting logic models can be used to: document accomplishments; organize data; prepare reports; and to define variance between the planned program and the actual program.

Logic Modeling Basics

The most basic logic model is a picture of why and how you believe your program; practice change or policy initiative will work. It uses words and/or pictures to describe the sequence of activities that you think will bring about change and how these activities are linked to the outcomes you expect your program, practice or policy initiative to achieve. It is important to remember that logic models present a theory about the expected program outcome. They do not demonstrate whether the program caused the observed outcome (you’ll need a specific kind of evaluation to do that - see Section Five and Six of this *Guide* for examples).

Logic model methodology helps you to describe, share, discuss, and improve program theory –in words and pictures – as you develop (plan, implement and evaluate) a program. Logic models depict the relationship between your planned work and your intended results.

In the context of training, a training program logic model is a graphic (a chain of events diagram) that describes your training program’s inputs, activities, outputs, outcomes and impacts, and outlines the anticipated linkages between each of these items. Logic models offer “if then” information that helps you think through training programming and evaluation.

- **Inputs** – The resources given to carry out training activities and to produce outputs and accomplish outcomes. Inputs include the human, financial, organizational and community resources that you have available to direct toward doing the work. Inputs are materials that you take in (e.g., staff, volunteers, facilities, equipment, practice standards, curricula, partners and money). A training program uses *inputs* to support training *activities*.
- **Activities** – Activities are what a program does with its inputs or resources - how it goes about transforming those inputs into products. Activities (or Strategies) are the types of services the program provides to fulfill its goals and objectives. In the training context, examples of activities would be training needs assessments, development of learning objectives, curriculum design meetings, training materials review, preparatory sessions with faculty, session workshops, exercises, breakout discussions, strategic planning, etc. Program *activities* or *strategies* result in *outputs*.

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

- **Outputs** – Outputs are the direct products of program activities. They are the measures of program process or implementation. Output data demonstrate the implementation of the program's activities. Output tells you what the program produced. Typical training outputs would include the number of participants trained, the number of training hours provided, number of curricula produced, and the number of training materials disseminated.
- **Outcomes** – Outcomes are the specific changes in program participants' behavior, knowledge, skills, and level of functioning. As a general rule of thumb, immediate outcomes are those expected at the conclusion of the training program. Short term outcomes should be attainable within 3-6 months of the training program, while longer term outcomes should be achievable within 1-2 years of a training program. Of course, timeframes for outcomes would be dependent on the specific outcome. Furthermore, outcomes should be logically tied to your theory of change – they should be relevant. Helpful words to consider when brainstorming training outcomes are: increase, decrease, maintain, reduce, improve, ensure ... ability to, skills for, knowledge of, confidence in, level of functioning in ...etc.

Table 1-9 Training Logic Model			
Statement of Problem What is your training program trying to address? What is your theory of change? What are the goals and objectives of your training?			
Training Inputs	→ Training Activities	→ Training Outputs	→ Training Outcomes
(Resources needed to design, operate and evaluate the training)	(Needs assessment, program design meetings, planning and logistics, material production, workshops, exercises, evaluation tasks)	(Number of sessions held, participants trained, materials distributed)	(Change in skill, knowledge, attitudes, behavior, policies, and improvements in communication, innovation, leadership, collaboration, safety, permanency, fairness, and well-being)

TOOLS AND RESOURCES – CHAPTER ONE

- Checklist to Determine if a Formal Instructional Design Approach is Needed
- Curriculum Design Worksheet
- Framework to Design a Training Plan
- Overview of Data Gathering Methods for Needs Assessment
- Personal Learning Styles Inventory
- Sample Questions to Obtain Training Participants' Needs
- Training Design Competencies Checklist
- Training Needs Assessment Sample Focus Group Protocol
- Training Needs Assessment Worksheet
- Training Planning Template
- Training Priorities Assessment Survey
- Training Program Logic Model Worksheet
- Training Program Logic Model
- Worksheet to Plan Evaluation Instruments
- Worksheet to Plan for Different Levels of Evaluation

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

CHAPTER ONE: References and Resources

- Altschuld, J.W., & Witkin, B.R. (2000). *From Needs Assessment to Action: Transforming Needs into Solutions Strategies*. Sage.
- American Educational Research Association www.aers.net
- American Evaluation Association www.eval.org
- American Psychological Association www.apa.org
- Arnold, W.E. & McClure, L. (1996). *Communication, Training and Development, 2nd Edition*. Waveland Press.
- Axon, D.A (2007). *Best Practices in Planning and Management Reporting, Second Edition: From Data to Decisions*. John C. Wiley and Sons.
- Beaudin, B. P., & Williams, R. E. (1990). "Improving human performance: Applying adult learning principles to enhance meetings," *Performance and Instruction*, 29(9), 7-11.
- Bloom B. S. (1956). *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York: David McKay Co., Inc.
- Boulmetis, J. & Dutwin, P. (2005). *The ABCs of Evaluation: Timeless Techniques for Program and Project Managers, 2nd Edition*. Jossey-Bass.
- Brookfield, S. (1992). "Why can't I get this right? Myths and realities in facilitating adult learning," *Adult Learning*, 3(6), pp. 12-15.
- Caffarella, R.S. (1994). *Planning Programs for Adult Learners: A Practical Guide for Educators, Trainers, and Staff Developers*. San Francisco, CA: Jossey-Bass.
- Center for Instructional Development and Research (2004). *Course Design*. University of Washington, available at <http://depts.washington.edu/cidrweb/CourseDesign.html>
- Chen, H.T. (1989). "Special Issue: The Theory-Driven Perspective," *Evaluation and Program Planning*, Vol. 12, pp. 297-398.
- Chen, H.T. (1990). *Theory-Driven Evaluations*. Sage.
- Cross, K. P. (1981). *Adults as Learners: Increasing Participation and Facilitating Learning*. San Francisco: Jossey-Bass.
- Darkenwald, G. G., & Merriam, S. B. (1982). *Adult Education: Foundations of Practice*. Harper and Row.
- Dick, W. & Cary, L. (1996). *The Systematic Design of Instruction (4th Ed.)*. New York: Harper Collins.
- Doran, G.T. (1981). "There's a S.M.A.R.T. way to write management's goals and objectives." *Management Review*, Volume 70, Issue 11.
- Flanagan, J.C. (1954). The Critical Incident Technique. *Psychological Bulletin*, Vol. 51 (4), p. 327-359.
- Fivars, G. (Ed) (1980). *Critical Incident Technique*, American Institutes for Research.
- Jung, C. G. (1971). *Psychological types (Collected works of C. G. Jung, volume 6)*. (3rd ed.). Princeton, NJ: Princeton University Press
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.
- Kemp, J. E. (1985). *The Instructional Design Process*. New York: Harper and Row.

CHAPTER ONE: TRAINING PROGRAM PLANNING AND EVALUATION

- Kirkpatrick, D.L. (1975). "Techniques for Evaluating Programs; Parts 1, 2, 3 and 4." In *Evaluating Training Programs*. ASTD.
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluation training programs: The four levels* (3^{ed}). San Francisco, Berrett- Koehler Publishers, Inc.
- Knowles, M.S. (1980). *The Modern Practice of Adult Education: From Pedagogy to Andragogy* (Rev. ed.). Chicago: Follett Publishing Co.
- Knowles, M. S. (1984) *Andragogy in Action. Applying modern principles of adult education*, San Francisco: Jossey Bass.
- Knowles, M.S. (1990). *The Adult Learner: A Neglected Species* (4th ed.). Houston: Gulf Publishing Co.
- Knowles, M.S. (1992). "Applying principles of adult learning in conference presentations," *Adult Learning*, 4(1), pp. 11-14.
- Knowles, M. S., Holton, E. G., & Swanson, R. A. (1998). *The Adult Learner: The Definitive Classic in Adult Education and Human Resources Development*. Houston, TX: Gulf Publishing Company.
- Knowles, M. S., Holton, E. F., & Swanson, R. A. (2005). *The Adult Learner* (6th ed.). Burlington, MA: Elsevier Butterworth-Heinemann.
- Kolb, D.A. (1984). *Experiential Learning*. Englewood Cliffs, NJ: Prentice Hall.
- Lawson, A.E. & Johnson, M. (2002). "The validity of Kolb learning styles," *Studies in Higher Education*, Vol. 27(1), pp. 79-90.
- Leshin, C.B., Pollock, J., & Reigeluth, C.M. (1992). *Instructional Design Strategies and Tactics*. Englewood Cliffs, NJ: Education Technology Publications.
- Lipsey, M.W. (1985). "Evaluation: The state of the art and the sorry state of the science." In D.S. Cordray (ed.), *Utilizing Prior Research in Evaluation Planning: New Directions for Program Evaluation*. Jossey-Bass.
- Pohl, Michael. (2000). *Learning to Think, Thinking to Learn: Models and Strategies to Develop a Classroom Culture of Thinking*. Cheltenham, Vic.: Hawker Brownlow.
- Ratliff, S.A. (1996). "Planning programs for adult learners," *Adult Education Quarterly*, 46(4), 240-242.
- Renate, M. & Caine, G. (1994). "Adult learning principles." In J. Helmich (ed.), *Making Connections: Teaching and the Human Brain*. Addison Wesley Publishing.
- Sternberg, R.J., Roediger, H.L., & Halpern, D.F. (2007). *Critical Thinking in Psychology*. Cambridge University Press.
- Thorndike, E.L. (1929). *Adult Learning*. New York: MacMillan.
- Weiss, C.H. (1998). *Evaluation: Methods for Studying Programs and Policies, Second Edition*. New Jersey: Prentice Hall.
- Witkin, B.R., & Altschuld, J.W. (1995). *Planning and Conducting Needs Assessments: A Practical Guide*. Newbury Park: Sage Publications

CHAPTER TWO

Training Satisfaction and Reaction Measurement

The assessment of satisfaction with the training process, program content, faculty, and venue is the most common type of training evaluation, with the focus on gathering participants' reactions to their training experience. As previously discussed, it is important to move beyond mere satisfaction measurement in dependency court training evaluation to include other levels of measurement – skill or learning acquisition, behavior or attitude change, and impact or outcome measurement. That being said, the assessment of training participants' reactions to, and satisfaction with, dependency court trainings should be included in any evaluation strategy as it provides critical information about the training process and implementation. In order to be useful, training satisfaction and reaction measurement must be tailored to the training's goals and learning objectives, and collected in a reliable and valid manner.

Chapter Two of this Guide provides recommendations and sample strategies for ensuring that the measurement of training participants' satisfaction is precise, valid, and provides useful information to training organizers. Some of the topics covered in this Chapter (e.g., the survey process, good questionnaire construction, building valid response categories, increasing response rates, the use of online measures, etc.) also have relevance to the assessment of learning, behavior and attitude change, and outcome measurement covered in other Sections. Much of what is presented in this Chapter on satisfaction and reaction measurement therefore, is foundational to all levels of good training evaluation practice.

SECTION 2.1

The Basics of Training Satisfaction and Reaction Measurement

Evaluating participants' satisfaction is an essential component of program evaluation. Measuring trainee satisfaction is critical to ensure the vitality of any training program. Unsatisfactory training programs will not meet the goals set forth by the organizers, the facilitators or the trainees in need of training, calling into question the utility of a given training program. This chapter focuses on the specific reasons why measuring satisfaction is important, and discusses the techniques required to measure satisfaction successfully. Because the primary methods of measuring satisfaction and reaction involve survey, interview and focus group procedures which are important tools for the other levels of measurement, this Section of the Guide should also be referenced when building learning acquisition, behavior change, and impact or outcome evaluation strategies.

Measuring participants' satisfaction allows training organizers to estimate how effective faculty was at engaging the audience. One way to ensure training has motivated trainees to learn is to measure their satisfaction with the program. If trainees are generally satisfied, it is an indication that they were willing to consider the curriculum being presented to them. If trainees are generally unsatisfied, it is an indication that barriers were present which prevented the trainees from considering the curriculum. To gather people's true impressions, it is imperative to allow people the opportunity to express their level of satisfaction anonymously. This is typically done by administering short questionnaires or surveys directly after a training program has concluded, or soon thereafter, without asking for respondents to provide their name.

Another reason it is important to measure satisfaction is to illustrate to trainees that their perspectives are important to training program organizers. Asking for feedback about the training demonstrates that you value participants' perspectives. The opportunity to express an opinion makes people feel as though they will have an influence on future training programs. Measuring satisfaction informs participants that training organizers and faculty want to do a good job and are interested in obtaining the perspective of trainees to ensure that they are doing a good job.

Measuring satisfaction also provides training organizers with quantitative information that can be used to establish standards of performance for a training program. If a training program is ongoing, it is important for organizers to know generally how satisfied participants are with the program. Determining the general satisfaction level allows organizers to determine whether trainings are continuing to meet the needs of trainees. It also allows organizers to determine whether specific faculty or training facilitators are equally as satisfying, or to determine whether specific training methods help or hinder a training program.

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

Satisfaction measurement may be used to address concerns and criticism voiced by individuals. Typically, people feel a need to express displeasure, an impulse not equally felt when something is pleasurable.³² In a training scenario, there are often a few vocal people who will readily express their concerns with some aspect of a training program, or perhaps their concerns with the entire training program itself. Measures of satisfaction can determine whether these feelings generalize to other trainees. In more instances than not, the majority of trainees do not share the criticism of others. By continuously and comprehensively measuring reactions to the training experience from as many participants as possible, training organizers will have a greater amount of information available to delineate valid from invalid criticisms.

To reduce the length of your satisfaction survey, consider a two phase satisfaction assessment process:

- 1) short, focused satisfaction feedback form given to all participants; and
- 2) follow-up in-depth focus group about training experience with a sample of participants.

In summary, measuring satisfaction is beneficial to both trainees and training organizers. For trainees, it gives them a voice to express themselves and their perceptions of the training, which includes them in the training process. For training organizers, measuring satisfaction provides general information regarding whether trainees received the training positively or negatively. The more areas of a training program that are measured the more information available to training organizers to decide how to improve on a training program.

Keep in Mind ...

While end-of course participant satisfaction questionnaires provide valuable information on participants' level of satisfaction with the training (their interest, their attitudes about the usefulness of the content, views on the quality of the training instructors and materials, etc.), satisfaction measurement does not provide valid information on workplace performance or the organizational impact of the training. In fact, most academic studies of satisfaction assessment or "reactionary" measurement have found that there is little correlation between participant satisfaction and learning or workplace performance results.³³ In other words, just because your evaluation found high trainee satisfaction with the course does not mean that trainees have actually learned or that they will apply what they have learned on the job, to change behavior or practice. You will need measurement at the other levels of evaluation to answer those questions.

General Guidelines for Evaluating Satisfaction

Determine What You Want to Find Out. Training organizers need to determine the key areas of the training in which they would like to measure satisfaction. There are several different components of a training that can be analyzed for participant satisfaction, including:

- The curriculum/material in general

³² Thagard, P. (2005).

³³ See for example: Barnett, S. M., & Ceci, S. J. (2002); Cree, V., & Macaulay, (2000); and Schunk, D. (2004).

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

- The organization of topics covered in the curriculum
- Whether participants' expectations for the training were met
- Facilitators/Presenters/Faculty
- Location of the training
- Atmosphere (facilities, meals, room temperature)

Of course, there is value to training organizers in obtaining reactions to every component of the training program, particularly with newer training programs. However, this may create a very cumbersome survey or extremely lengthy feedback form and decrease the likelihood that participants will complete the form or respond thoughtfully. Instead, training organizers should consider a more focused approach to satisfaction measurement for both new and more established training programs – examining the specific components of the training program for which they feel information about satisfaction would be most critical. Or, training organizers can consider a two-phase process: 1) provide a shorter, more focused satisfaction feedback form to all participants; and then 2) follow-up with a sample of training participants and engage them in a focused, more in-depth discussion about their training experience (e.g., via a focus group or interview process, or via a more detailed survey).

Quantify Participants' Responses. When designing participant reaction or feedback forms, it is helpful to design an instrument that will *quantify* participants' reactions. Quantifying satisfaction responses simplifies data analysis. Turning people's responses into numbers increases training organizers' ability (and other people exposed to the data) to understand and discuss what is going on in a training program. Quantifying responses is generally accomplished in surveys using measurement scales. For example, a Likert scale, which is a rating continuum anchored by dichotomous concepts, such as 'agree' or 'disagree,' is a common method for quantifying responses. The number of intervals on the scale can vary, but research finds that people map their perceptions best onto 5 or 7 point scales³⁴ (See subsequent Sections of this Chapter for more detailed information on the use of measurement scales). With respect to quantification, we are not advocating abandoning open-ended questions (see below). However, we do recommend limiting them, as drawing out qualitative themes during data analysis is extremely time consuming if there are a large amount of respondents. The ideal evaluation form provides the maximum amount of information and requires the minimum amount of time.

³⁴ Schwarz & Sudman (1996).

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

Encourage Open Responses. While quantification simplifies data collection and analysis, turning responses into numbers does not allow trainees to express any personal perspectives not accounted for on a questionnaire, survey, or interview. Qualitative responses allow participants the opportunity to share their perceptions, allowing participants' the opportunity to bring new ideas to the attention of the training organizers – ideas or concerns that organizers may not have been aware of when creating the evaluation form. Therefore, it is always optimal to include some open-ended questions on basic evaluation forms to encourage participants to share any comments or suggestions. Additional open-ended questions should be included and tailored to any area of a training program where organizers believe it is important in order to gauge trainees' perspectives. In addition, opportunities for respondents to clarify their responses are critical to your understanding of the data. For example, if respondents report that they were not satisfied with their training experience but they are not given an opportunity to tell you why they felt that way, training organizers may not have sufficient data to make appropriate program modifications based on this feedback. Be sure to include opportunities for respondents to comment or further explain their responses, in order to build an explanatory model for your survey findings -the "why's" behind the satisfaction judgments.

Tip from Field-Test Sites: Some of the sites participating in field tests of the recommendations and tools in this Guide provided training participants with a short version of the evaluation form at the beginning of the training program. Participants were encouraged to use the form to record their immediate thoughts, impressions, ideas, etc. while they were still fresh in their minds. At the conclusion of the training, participants were directed to complete a full online version of the evaluation form. Participants could use the short form to refresh their memories if needed, but training organizers benefited from the ease of data collection and analysis afforded by the use of a web-based online survey.

Encourage Immediate Responses to Ensure Good Response Rates. Getting participants to complete evaluation forms as part of the training program allows organizers to measure the perceptions of the participants while the information is still fresh in their minds. It also ensures that most of the participants' perspectives will be accounted for, by increasing the likelihood that forms will be completed and returned. Alternatively, training organizers may ask participants to complete and return evaluation forms at a later date, or may have participants complete an on-line survey at a later date. If this method is selected, evaluation forms should be returned within two weeks of concluding the training (organizers should follow-up with reminder emails to encourage completion and return of the evaluation forms at one week to ten days). The longer you wait to assess satisfaction the more likely that response rates will decrease.³⁵ In addition, the accuracy of satisfaction measures decreases the further away in time the evaluation is conducted from the completion of the training.³⁶ Even if evaluators are planning on conducting a follow-up survey 3 – 6 months after the training (i.e., to measure learning retention or behavior change), an initial evaluation of satisfaction should be assessed

³⁵ Fowler, F.J. (1988); Heerwegh, D. (2005).

³⁶ *Supra* note 35.

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

as soon as possible after the training session. Tips for increasing survey response rates are covered in later chapters of this Section and in the Tools and Resources for this Section.

Encourage Honest Responses. Clearly, it is important to obtain honest responses from participants about their training experience. If you provide respondents with information about the goals of the survey and how critical their feedback is to the improvement of future training programs, you can motivate respondents to provide honest feedback. But the best way to ensure that participants' satisfaction responses are genuine is to ensure their anonymity or confidentiality. This can be accomplished by collecting data in a manner which does not tie the information provided to the person supplying the information. For example, don't have participants put their names on questionnaires or surveys (you can assign a code to the completed survey if you need to track their responses for pre- and post-test comparisons). Have interviewers erase any identifiable information that may have been collected during interviews (i.e., name and contact phone number) before handing over the information for data analysis. However, even though you have implemented these procedures, if training participants aren't aware of them they may still feel that they can't provide open and honest responses. Be sure to include a statement on evaluation feedback forms that lets training participants know that their responses will be anonymous (e.g., "Your name will not be associated with the responses you provide"). Another way to ensure anonymity is to report the findings as aggregate numbers, and abstain from presenting data in a way that may allow people to identify someone with a particular response. For example, you can help maintain anonymity by reporting findings by the role of respondent. But be careful ... if there was only one or two judges attending the training, don't report findings as "judicial officers attending the training felt that ..." as this may identify the individual respondents. If it is critical to supply information that can be tied to an individual for your evaluation report, be sure to ask that individual if you may report that finding, quote them, or otherwise refer to their feedback. Ensuring participants of their anonymity, and taking lengths to ensure their responses are confidential, are the best ways to get an honest measure of trainee satisfaction.

Develop Acceptable Standards for Trainings. It is important to develop criteria of acceptable standards for your training program. Establish a baseline of how you would expect trainees to respond to a training program – for example, consider an acceptable average satisfaction level that you would expect from trainees (e.g., "at least 80% of trainees will express satisfaction with the training materials"). Knowing the acceptable average level of satisfaction among trainees allows organizers to determine what changes are necessary to maintain or improve trainee satisfaction. In addition to developing your own acceptable standards for training, be sure to determine whether training funders or other relevant administrative bodies have established standards that can guide your understanding of training satisfaction.

Compare Current Participants' Responses to Standard. Measuring satisfaction at each training allows organizers to generate not only a baseline average from which to compare future trainings, but also allows organizers to determine what standards they feel are acceptable for a given training. Once realistic standards have been established, you should continue to evaluate trainee satisfaction with various components of the

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

training program and compare your findings with the standards. If the standards are not met, make a change and modify the training program accordingly.

Communicate Reactions as Appropriate. Finally, communicate the results to the appropriate audience. Typically the appropriate audience for training evaluation results includes the training organizer, facilitator, faculty, training advisory committee, and training funders. In some instances, other parties may be interested in the results. You may want to make an abbreviated report of the training evaluation results available to your training participants -doing so provides them with an opportunity to see how their feedback was used and may serve to encourage participation in your future or follow-up evaluation activities. To determine exactly how you should communicate trainee satisfaction, consider who wants to see the results, and whether it is appropriate for them to do so. See Section Seven of the Guide for recommendations for analysis and reporting of training evaluation findings.

Consideration of the guidelines for evaluating satisfaction above can help training organizers determine the level of inquiry desired out of a satisfaction measure. Knowing what you want to find out, how you want to find out about it (i.e., quantitative or qualitative), and how the information will be used provides organizers a general framework for developing their evaluation. If the goal is to measure satisfaction on only a few components of a training program, the questionnaire can be focused and brief. If the training is new, or organizers are interested in every component of the training, then a more in-depth survey with greater detail may be required.

Types of Questions to Ask When Evaluating Satisfaction

Clearly training organizers can ask satisfaction-based questions about almost any aspect of a training program. As a result, it's helpful to group satisfaction questions into "types" or categories of questions and then to prioritize among those questions. The following text provides some sample questions under each general category or domain of training satisfaction measurement. Many of the sample questions below also lend themselves to follow-up, open-ended questions should training organizers require more detailed explanation for a specific response. The list of sample questions is illustrative only and is not meant to be exhaustive of the satisfaction questions that might be asked under each category. See later Chapters of this Section for more detailed instruction about question construction.

General Satisfaction Questions

General satisfaction questions are designed to capture trainees' overall impression of the training. Some sample general satisfaction questions are provided below. Unless otherwise noted, these sample questions would include a rating scale for responses from "Highly Satisfied" to "Not at all Satisfied" – more about the use of rating scales and determining appropriate anchors for those scales in later chapters of this Section.

- How satisfied are you with the overall training?

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

- How would you rate the training program overall? [Provide rating scales from “Very Useful” to “Not at all Useful,” and “Very Interesting” to “Not at all Interesting”]
- How satisfied are you that the time you spent at this training is beneficial to your job?
- To what extent do you feel your personal learning objectives have been achieved? [Provide rating scale from “Fully” to “Not at All”]
- To what extent do you feel your expectations for this training have been met? [Provide rating scale from “Fully” to “Not at All”]
- Which of your personal learning objectives were not achieved at this training and why?
- Which of your expectations for the training were not met and why?

Specific Questions about the Training Curriculum

An important component of trainee satisfaction involves gauging impressions of the training material. If trainees are not stimulated by the material, they will not incorporate it into practice. Therefore, training evaluations should assess trainee satisfaction with the material presented during training. Examples include:

- How satisfied were you with the way the training program was organized (i.e., the order of topics)?
- Which parts of the training curriculum will be most useful to you in your job?
- Which parts of the training curriculum will be least useful, or not at all useful, to you in your job?
- Please indicate your agreement with the following statement: The materials provided to me at this training were tailored to meet my training needs. [Provide a rating scale from “Totally Agree” to “Totally Disagree”].
- Are there any subjects you would like to have seen included that were not included?
- To make way for additional curriculum materials, what would you omit from the training program?

Specific Questions about Presenters and Presentations

A key component of trainee satisfaction involves impressions of the presenters or faculty. Measuring satisfaction with the presenter provides an indication as to whether the presenter was successful in establishing his or her credibility and holding the audience. Presenting information in a clear and concise way that is easy to follow is essential for facilitating adult learning. Training evaluations should incorporate items investigating trainee satisfaction with both the presenters and the presentations. Examples include:

- How satisfied were you with the presenter overall? [Insert specific presenter or presentation for these questions as appropriate]
- How satisfied were you with the presenter’s style?
- How satisfied were you with the way the subject matter was presented?
- How satisfied were you with the pace of the presentation?
- How satisfied were you with the organization of the presentation?

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

- Please rate the presenter's ability to engage the audience [Provide rating scale from "Highly Skilled" to "Not at all Skilled"]
- How useful was the exercise to you as a means of illustrating the application of concepts? [Provide rating scale from "Very Useful" to "Not at all Useful"]
- How satisfied were you with the materials handed out to participants?
- Please rate the usefulness of the materials handed out to participants [Provide rating scale from "Very Useful" to "Not at all Useful"]

Specific Questions about Facilities

It's often valuable to obtain the reactions of training participants to matters outside of the evaluation of the learning itself (e.g., training facilities, training administration, etc.). By using a well-constructed and effective feedback form – not one that is skewed to prompt favorable comments – useful data can be obtained to help plan future training. Clearly, the training environment or atmosphere may also impact trainee satisfaction levels. Trainees will not be motivated to listen if they are not comfortable with the facilities. To assess satisfaction with training environments, questions should be included about satisfaction with the location of training, the hotel where trainees stayed, the rooms in which the training occurred, as well as the food that was served. Examples include:

- How satisfied were you with the location of the training?
- How satisfied were you with the room in which the training was held?
- How satisfied were you with the sound quality in the room in which the training was held?
- How satisfied were you with the food served during the training?
- How would you rate the quality of the food served during the training? [Provide rating scale from "Excellent" to Poor"]
- How would you rate the comfort of the training room? [Provide rating scale from "Excellent" to "Poor"]
- How would you rate the ease of travel to the training location? [Provide rating scale from "Excellent" to "Poor"]

In summary, measuring satisfaction is important to ensure that the audience is accepting of the training - providing organizers with some evidence that the program was able to overcome the barriers associated with adult learning mentioned in Section Two of this *Guide*. Measuring satisfaction also provides training organizers with an estimate of how well the training was received.

See the Chapter Two Tools and Resources for additional sample satisfaction measurement questions and sample or template satisfaction measurement forms. See also the Question Bank Tool included with the Tools and Resources materials for more sample satisfaction and reaction measurement questions.

SECTION 2.2

The Survey Process

Clearly a major tool in the assessment of satisfaction with a training program is the survey or questionnaire. However, in discussions on surveying generally, and on satisfaction or reaction measurement specifically, the focus is often incorrectly placed only on the survey instrument and not on the entire survey process. The entire survey process includes defining the survey objectives, developing a sample frame (a list of the population of individuals of interest from which a sample to receive the survey will be drawn), designing questions, specifying the strategy for data collection, and conducting the appropriate analyses. This entire process is important to achieving acceptable response rates, obtaining reliable and valid findings from your survey, and receiving acceptance for the recommendations that are generated from your results.

The Formal Survey Process³⁷

1. Define the Survey Objectives:
 - a. Specify the population of interest (who you need to survey)
 - b. Determine the type of data to be collected
 - c. Determine the desired precision of the results
2. Determine who will be sampled:
 - a. Specify the method of sample selection as either probability-based or convenience-based.³⁸ For training evaluations, this is almost always the entire group of individuals who attended the training program
 - b. Create a sampling frame if necessary (a list of individuals from whom you will draw the survey sample)
 - c. Select the sample
3. Create and test the survey instrument:
 - a. Choose the response mode (handed out at training, mailed, web-based)
 - b. Draft the questions
 - c. Pre-test and revise the survey instrument
4. Contact respondents throughout the survey process
 - a. Notify respondents that the survey is coming
 - b. Provide instructions for completing and returning the survey at the delivery stage
 - c. Remind respondents to complete the survey through post-delivery
 - d. Send post-delivery thank-you's for completing the survey
 - e. Conduct non-response follow-up for those who did not return the survey
5. Collect data, data reduction, and data analysis

³⁷ The Total Design Method, Dillman (1978; 2000); see also Fowler (1988).

³⁸ In a probability sample, the probability with which an individual is selected into a sample can be computed. When the probability cannot be computed, the sample is called a convenience sample (it is more "convenient" to not have to worry about the ability to compute probabilities of selection).

Important Considerations in Planning and Designing a Survey

Increasing Survey Response Rates

The response rate is used to evaluate the success of your data collection effort. It is simply the number of people responding (or interviewed) divided by the total number of all of the people sampled. The denominator in this equation includes all of the people who were selected to receive a survey (e.g., attended the training) but did not respond for whatever reason. With respect to surveys in training evaluations, your response rate is calculated by taking the total number of completed and returned surveys and dividing that number by the total number of training participants (those individuals who received the survey in the first place).

Clearly, the greater the response rate, the more confidence you can have that evaluation findings derived from the survey represent the experiences of your training participants. If your response rate is low, then your findings may be peculiar to those individuals who were more inclined to complete a survey - and not truly reflective of the experience of the majority of training participants. Response rates vary widely for different types of surveys. Customer satisfaction surveys and market research surveys often have response rates in the 10% - 30% range.³⁹ Employee surveys typically have a response rate of 25% - 60%.⁴⁰ An important incentive to survey respondents is that their opinions will be heard and action will be taken based on their feedback. If respondents believe that participating in a survey will result in real improvements, response rates may increase, as will the quality of the feedback. Response rates can soar past 85% (about 43 responses for every 50 invitations sent) when the respondent population is motivated and the survey is well-executed.⁴¹ Response rates can also fall below 2% (about 1 response for every 50 invitations sent) when the respondent population is less-targeted, when contact information is unreliable, or where there is less incentive or little motivation to respond. Regardless of the type of survey you are conducting, you can have a major effect on the number of respondents who complete your survey. In order to increase your response rates, attention needs to be paid both at the survey design stage and at the collection stage.

To motivate respondents to agree to answer a survey and to complete it accurately, survey research demonstrates that the length of a survey is not as important as its design.⁴² Follow the strategies listed below to help increase your survey response rates:

- Particularly important when a respondent first views a survey is that it looks easy to do and the instructions are clear and consistent. To format a survey to maximize responses, follow some simple principles that give visual signals that the survey is easy to complete, such as creating a sense of space in the survey by eliminating the unnecessary use of lines - a long survey with lots of white space looks easier to complete than a two page survey filled with lots of print and lines.

³⁹ See for example, Herberlien & Baumgartner, 1978; Steele, Schwendig & Kilpatrick, 1992; Yammarino, Skinner & Childers, 1991

⁴⁰ *Ibid.*

⁴¹ *Ibid.*

⁴² Salant, P. & Dillman, D.A. (1994).

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

- While thinking about including space in your survey's design, be sure to make your survey as professional-looking as possible. This will increase your credibility and help to ensure that people invited to take the survey believe that their responses are important and will be used.
- Use the logic model, objectives and purpose for the evaluation that you developed to base your evaluation questions on - that is a useful way to cut back on extraneous evaluation questions. Focus your questions on evaluating the pieces of your training program that are most important to you and to your stakeholders.
- If your survey is not completed immediately after the training, don't give participants too much time to complete the survey. While the measurement of the application of training content and resulting behavior and practice change benefits from delay, response rates do not. With satisfaction measurement you also want respondents to be able to recall their impressions of the training experience accurately - this is facilitated by short turn-around times between the completion of training and satisfaction measurement. We recommend 2 weeks as a run time for training surveys in which it is important to get a full response (as measured from sending out the survey to closing the survey for analysis).
- Be sure to use follow-up reminders to non-completing survey recipients after the original invitation is sent. A follow-up within 10 days after the initial invitation is optimal.
- Offer an incentive to complete and return the survey. Research results demonstrate that incentives will typically increase response rates by 10-15% (depending on the quality and attractiveness of the incentive to your target audience).⁴³
- Personalize the survey - invite respondents to take surveys that are sent to them after the training by addressing the invitation with their name. Research has shown that personalization of e-mailed survey invites ('Dear Kathy' instead of 'Dear Training Participant') can increase response rates by 8% or more.⁴⁴

Tips from Field-Test Sites

In order to increase survey response rates, some of the sites testing the recommendations and strategies in this Guide offered additional technical assistance and resource materials for those completing surveys; opportunities to serve as faculty or make a presentation at future trainings; opportunities to include more team members in future trainings (where registration had been limited in the past); and including a letter from the Chief Judge in the invitation to complete the survey (indicating the value of respondents' feedback to future training programs and court improvement).

⁴³ Herberlien & Baumgartner, 1978; Steele, Schwendig & Kilpatrick, 1992; Yammarino, Skinner & Childers, 1991.

⁴⁴ Heergweth (2005).

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

- With online surveys, give consideration to the time you send your survey link and invitation out. Research indicates that response rates and times are best for surveys sent out between 6:00AM and 9:00 AM, at the beginning of the work day – but not on Monday morning.⁴⁵
- Make sure you make the link between the survey and the needs of the individuals who you have invited to take the survey - make the survey salient. Reinforce for training participants (during the training itself and during the survey invitation) how the evaluation findings are used to design and implement trainings that assist them in performing their job functions and achieving their court improvement goals. Salience of an issue to the population receiving the survey has been found to have a strong positive correlation with response rate for mail and web-based surveys.⁴⁶ The greater the importance of the topic to the individual the more likely they are to complete and return the survey. If individuals attach little interest or importance to the content of the survey, then it won't matter if the survey is short and appears to be easy; the person is still unlikely to respond.⁴⁷

Question Order

Question sequence requires careful thought to reduce the likelihood of bias in surveys, interviews or focus groups. Order can affect results when a general question and a more specific question are asked in sequence. Research using “split-ballot experiments,” in which the order of questions is rotated, suggests that results from a general question are likely to be affected depending on whether it comes before or after a specific question.

⁴⁸ For example, consider the order of questions in this hypothetical training needs assessment:

[A Specific Question] How would you describe the quality of child abuse and neglect permanency hearings in your jurisdiction? (Please circle the appropriate number below):
1. Very Poor 2. Poor 3. Good 4. Very Good 5. Don't Know

[General Question] Overall, how would you describe the quality of child abuse and neglect hearing practice in your jurisdiction? (Please circle the appropriate number below):
1. Very Poor 2. Poor 3. Good 4. Very Good 5. Don't Know

In the examples above, placing the specific question first may bias respondents' answers to the more general question that follows because their thoughts about the permanency hearing have become salient as they consider their “overall” rating of hearing practice.

Clearly, ensuring that the order of your questions doesn't potentially bias responses is an important consideration when putting together the questions for your survey. Also important, is a logical sequence for the questions. Questions tapping the same domain, for example, are best grouped together (e.g., questions about specific faculty and workshops should be grouped together, as should questions about training logistics, venue, etc.). Feedback about the logical order for questions, ease of understanding and navigation through the survey, should also be obtained before finalizing the survey instrument (see pre-testing below).

⁴⁵ *Ibid.*

⁴⁶ Sheehan & McMillan (1999); Watt (1999).

⁴⁷ Martin (1994).

⁴⁸ Krosnick, J.A. (1999).

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

Pre-Testing and Piloting Evaluation Questions

Before finalizing your survey, it's important to pre-test or pilot the survey instrument. Pre-testing or piloting is an important component of the evaluation process. It affords researchers the opportunity to identify any problems with instrumentation before beginning formal data collection. Ideally, your evaluation instrumentation pre-test or piloting procedures should involve having your survey reviewed by experts who are knowledgeable about survey question construction and the subject matter, as well knowledgeable about the objectives of the training program itself. Experts can tell you which questions appear too complex to be administered easily and which are too long or too difficult to be answered accurately. Experts can also help you prioritize among questions in order to ensure your final survey instrument is concise and lacking in superfluous questions – that every question is logically tied to the objectives of the training and will provide useful information to training organizers. Reach out to your evaluation research community, local universities or the National Child Welfare Resource Center on Legal and Judicial Issues for technical assistance with survey construction and training evaluation instrument design.

Pre-testing or piloting of your instrumentation should also involve having instruments reviewed by potential respondents. In the training context, potential respondents are people who are eligible to be part of the survey sample – the kind of people you want to hear from. For example, if your training is focused on judges, then a review of your instrument by judges who will not be involved in the training would be useful. A review by potential respondents helps to guarantee that the survey's questions are understood by your target audience, and are meaningful and inclusive of all important ideas.

In addition to a review of the survey content, be sure to pre-test the time it takes to complete survey instruments. This is especially important when you are suggesting timeframes to your respondents for the completion of surveys. Have a sample of individuals who are unfamiliar with the survey (i.e., have not been involved in the survey's design) take the survey and time their completion. This will provide you with a fairly accurate estimate of time that respondents can expect it to take to complete the survey. You can then cite this time in your invitations to take the survey (web-based or paper forms).

SECTION 2.3

Question Construction

Above all, the questions you include on any evaluation instrument should be purposeful. Purposeful questions are those that are logically related to your objectives. Construction of your training evaluation questionnaire, whether it is primarily designed to gauge satisfaction with the training experience or is more comprehensively aimed (as this *Guide* recommends) to assess satisfaction, learning, and behavior change, begins with an identification of the following:

- How will the survey or questionnaire fit with other training evaluation data collection methods?
- What kinds of questions need to be asked, based on the training needs assessment, identification of learning objectives and training program logic model?
- Who will receive the questionnaire and when in the training and evaluation process will responses be sought?
- What types of decisions will be made from the collected data?
- What kinds of questions or items will elicit useable data?
- Are there existing questions that have been used successfully in other surveys that we can adapt?
- What format will be used for the questionnaire – keeping in mind that the format should be as easy as possible for the user and for analysis?
- How will the data be analyzed and collated with other evaluation data to establish training priorities and to make determinations about training impacts and outcomes?

Types of Questions - Closed vs. Open-Ended Questioning

There are two types of survey questions: open-ended questions and closed-ended questions. Each type has advantages and disadvantages. Closed-ended questions provide a range of answer categories or options from which participants can choose. Open-ended questions provide no answer categories but allow participants to express ideas in the language they choose. Closed-ended questions are considered more efficient for both the respondent and for analysis. Respondents experience fewer burdens in completing closed-ended questions than open-ended ones, thus insuring you a greater response and less missing data. Additionally, you can tally questions with answer categories more quickly. However, open-ended questions provide respondents with the opportunity to tell you more about their opinion or experiences, and to clarify close-ended responses (e.g., by providing answers to “please explain” or “why do you feel that way?”). Open-ended questions provide the necessary explanatory models for quantitative survey findings.

Clearly, question types should depend largely on the needs and goals of the training, the length of survey or questionnaire, and the type of analyses planned. All types of questions and responses can be correct, if used

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

appropriately. Three types of questions are typically used to measure satisfaction and reaction to dependency court trainings: closed-ended scale item questions, closed-ended forced choice (i.e., true/false) questions, and open-ended questions.

Closed-Ended – Forced-choice or closed-ended question formats ask respondents to select one of two options – typically a *true* or *false* or *yes* or *no*. Forced choice response questions are often used when determining *if*'s – if training was

Example of Forced Choice:

Was the information provided today useful?

Yes

No

informative, or if the materials were clear. Forced choice (i.e., yes or no) questions do not allow for variation in responses and are only preferable when there should be no variation in response. Because individuals are forced to select between two options (such as yes or no), evaluators have no way of knowing why the response was selected. For example, if a respondent indicated that the information provided was not useful, without a follow-up question evaluators would not know why the respondent felt that way. This means the information could have been highly informative or only slightly informative. Other forced-choice question formats include rankings, paired comparisons, and 'most' or 'least' important questions.

- *Rankings* – Items are ranked by respondents in order of importance or preference, with only one rank assigned per item. This format works best with 15 items or fewer.
- *Paired Comparisons* – Items are presented in pairs, and respondents choose one from each pair.
- *Most and Least Important* – Respondents choose the three to five most important (or most desirable) and the three to five least important (or least desirable) from a list of statements.

Open-Ended – Open-ended questions allow respondents to give answers in their own words. These questions are useful if you are interested in getting unanticipated answers or in learning about the world as your respondents really see it. Also, some respondents prefer to state their views in their own words, which sometimes results in quotable material that can be useful for your evaluation report. The responses to open-ended questions, however, are often more difficult to compare and interpret.

Table 2-1 When to Use Open- and Closed-Ended Questions		
	If Yes, Use Open-Ended Questions	If Yes, Use Closed-Ended Questions
Purpose	Respondents own words are essential (to give voice, to obtain quotes).	You want data that are rated or ranked and you have a good idea of how to order the ratings (or response categories) in advance
Respondent Characteristics	Respondents are capable of providing answers in their own words. Respondents are willing to provide answers in their own words.	You want respondents to answer using a pre-specified set of response choices to facilitate interpretation and comparison across respondents.
Asking the Question	Potential choices or possible response categories are unknown.	You prefer that respondents choose among known choices.
Analyzing the Results	You have the skills to analyze respondents' comments even though answers may vary considerably.	You prefer to quantify or count the number of responses per specified choice or category.
Reporting the Results	You will provide individual or grouped verbal responses.	You will report statistical data (e.g., percentages, frequencies, average ratings).

Pitfalls to Avoid When Designing Questions

Conversational Language – A survey is not a conversation. To get accurate information, survey questions rely on standard grammar, punctuation and spelling. You should use words that maximize understanding for everyone involved in the survey. All questions should be reviewed and tested by people who are proficient in reading and speaking the language in which the survey is written, by content experts, and ideally, by potential respondents. Complete sentences should be used, whether as statements or questions, expressing clear and complete thoughts. In addition, avoid using slang and colloquialisms because they appear unprofessional, go out of fashion quickly, and may not be familiar to all of your survey respondents. However, it may be appropriate to use phrases that are “terms of art” or are familiar to your respondents if they come from a homogeneous group (e.g., individuals who share a common professional language).

Poor – Stakeholder role?

Better – Please identify your role in the dependency court system in the space provided (e.g., judge, child welfare worker, etc.).

Poor – Any thoughts on curriculum content?

Better – Please use the space provided below to include any thoughts you might have for improving the training curriculum content.

Complete Sentences

– an over-reliance on conversational language and incomplete sentences makes your survey appear unprofessional and also may introduce confusion in your respondents about what you are asking or requiring of them.

Abbreviations and Acronyms – Avoid using abbreviations and acronyms in questions unless you are sure that they are commonly understood. Most people are familiar with

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

ASFA and CFSR, for example, in the dependency court system context. If in any doubt, it's important to spell out abbreviations and acronyms.

Jargon and Technical Expressions – It is best to avoid the use of jargon and technical terms in your questions unless you have good reason to believe that all of your respondents are familiar with the terms.

Ambiguous Questions – Concrete questions are precise and unambiguous. Questions may be defined as precise and unambiguous when, without prompting, two or more potential respondents agree on the meaning of the words used in the question. In the training context, for example, words like “quality” may be difficult to interpret as individual respondents may have different perceptions or definitions of “quality.” The more detail you can provide in a question, the more reliable the answer is likely to be. To help make questions more precise and concrete, you may also consider whether adding a time frame might be appropriate.

Use the following scale to rate each question: **Poor 1 2 3 4 5 Excellent**

Ambiguous – *Quality* of written material provided 1 2 3 4 5

Better – *Relevance* of written material provided to my role 1 2 3 4 5

Less precise – Will you apply the training learned?

More precise – Will you apply the training learned in your job in the next month?

Ambiguous wording – Use of words like *quality* of material may be difficult to interpret by respondents and lead to difficulty in interpreting responses

Biasing Words, Phrases and Response Sets – Biasing words or phrases elicit emotional responses that may have little to do with the issues addressed by the survey. They are considered biasing because they trigger emotional responses or prejudice. Bias may also arise in your survey if you do not fully understand the culture and values of the respondents and ask questions that are inadvertently offensive. To guard against this possibility, you need to have all questions reviewed and pilot-tested before you use them. In addition, bias may enter into your survey through your choice of response sets – a biased response set is created when the response categories offered do not include all relevant options. The omission of the extreme on the negative end of a rating scale, for example, biases responses

Biased – In your opinion, did the training do a good job of getting judges to care more about timely permanency for children?

Better – In your opinion, did the training do a good job of focusing judges on the importance of timely permanency for children?

Biased – Do you believe that the new protocol introduced at this training will finally get judges to exercise their oversight role?

Better – Do you believe that the new protocol introduced at this training will assist judges in exercising their oversight role?

Biased – Overall, how would you rate the presentation on timely permanency?
Excellent Good Fair Poor

Better – Overall, how would you rate the presentation on timely permanency?
Excellent Good Fair Poor

Biased Response Set – The responses do not include an option for something worse than fair, such as poor, and are thus skewed toward a positive response

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

towards favorable ratings.

Double-barreled Questions - A double-barreled question contains two separate ideas. An example is “Do you think we should continue to train on ASFA and strategies for effective court-agency collaboration?” This question is really twofold: “Do you think we should continue to train on ASFA?” and “Do you think we should continue to train on strategies for effective court-agency collaboration?” Some respondents may endorse continued education on ASFA, some may endorse training on collaboration, some may endorse one but not the other, and some may choose not to endorse either type of training. No matter what the respondent answers to a double-barreled question, however, you will not know exactly what he or she means. To avoid asking double-barreled questions, check the use of the word “and” in your questions.

Double-barreled – Please indicate your agreement with the following statement on a scale from “5” Strongly Agree to “1” Strongly Disagree

Poor - Overall, the presenter was prepared and organized 5 4 3 2 1

Better - Overall the presenter was prepared 5 4 3 2 1

Overall the presenter was organized 5 4 3 2 1

Double-barreled– two separate ideas are contained within one question -respondents may feel one way about one idea and another way about the other.

Negative Questions - Negative questions are difficult for many respondents to answer because they require an exercise in logical thinking. For example, suppose a question asks respondents if they agree or disagree with the statement “Prior to the adjudication stage, dependency mediation should not be used.” Some respondents will fail to read the word “not.” Others will mistakenly translate the negative into the positive and believe the question is “Do I think mediation should be used prior to adjudication in dependency cases?” If you use a negative question, be sure to emphasize the negative word: “Prior to the adjudication stage, dependency mediation should not be used.” Be careful not to use double negative wording as well, such as “Do you disagree with those that do not want to expand the mediation program?” Double negatives are not only grammatically incorrect but they are also very confusing.

Negative Question– Do you agree with the statement that prior to the adjudication stage, dependency mediation should not be used?

Better - Do you agree with the statement that dependency mediation should not be used prior to adjudication?

Better - Do you agree with the statement that dependency mediation should only be used after adjudication?

Negative Questions
- are often confusing for respondents

Measurement Choice: Choosing Response Categories

Response choices, or the choice given to respondents from among which they select their answers, take several forms. Deciding which kind of response choices you should use first involves determining what type of information you need. A categorical or nominal response choice involves categories, such as male and female, yes and no, ‘applies’ or ‘does not apply.’ These types of responses have no numerical or preferential values – they are simply correct or incorrect, true or false. A second type of response choice is ordinal, in which respondents are asked to rate or order the items in a list (e.g., from very positive to very negative). Numerical response choices call for numbers, such as years of experience or number of trainings attended.

Rating Scales - By far the most common types of questions for training evaluation are closed-ended rating scale items (e.g., Likert scales).⁴⁹ Rating scales are usually 5 point scales with descriptions (anchors) at each end of

Example of a Rating Scale: Please indicate your degree of agreement with the following statement, by circling the appropriate number on a scale from “5” Strongly Agree to “1” Strongly Disagree.

The presentation of information was easy to understand.

Strongly Agree 5 4 3 2 1 Strongly Disagree

the scale. Questions may be asked about frequency (never – always), amount (least – most) or satisfaction (not at all – very). Alternatively, you can use adjectives for each point of the scale (never, occasionally, some of the time, frequently, always). While 5 point scales are the most common, 4 point or 6 point scales can be used to prevent artificial clustering around a midpoint. Rating scales may also determine the strength or intensity of judgment

regarding satisfaction with training delivery, course content, or other features or processes involved in the training program implementation. Some research indicated that the anchors of scales should be alternated (e.g., leading with the positive anchor for some sets of questions and then leading with the negative anchor for other sets of questions) in order to prevent a “response-pattern” or the tendency for respondents to go down a list of questions, not pay attention to those questions and mechanically circling scores.⁵⁰

Poor balance – Yes, constantly; Yes very often; Yes, once; No, never.

Better balance –Yes, constantly; Yes, very often; Yes, fairly often; Yes, a couple of times; Yes, once; No, never.

Guidelines for Determining Ordered Responses, Scales or Rankings

- **Use a meaningful scale.** A meaningful scale is one that makes sense in terms of the survey’s specific objectives. To choose among potential scales, you may test one or more scales on a preliminary or pre-test basis and select the one that gives a good “spread” of answers and is the most meaningful to respondents.

⁴⁹ Likert scales are probably the most widely used response scale featured in surveys –Created by Rensis Likert in the 1930s, his original scale featured five points.

⁵⁰ Fowler, F.J. (1988); Fowler, F.J. (2001).

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

- **Balance all responses.** A scale is balanced when the two endpoints mean the opposite of each other and the intervals between the points on the scale are about equal. “Much worse” is the opposite of “much better,” and the meaning of the interval between “much worse” and “somewhat worse” is similar in degree to the interval between “somewhat better” and “much better.”
- **Use a neutral response category only if it is valid.** Provide a neutral category only when you are sure it is a valid response. A neutral category is either a middle point (“neither satisfied nor dissatisfied”) or a “no opinion” or “neutral” or “don’t know” option. Some survey research has indicated that providing neutral choices gives respondents an excuse for not answering questions.⁵¹ If you think your respondents might react this way to neutral choices, pre-test your questions with and without neutral choices and compare the results. How many responses cluster around the middle? Alternatively, some survey researchers suggest that respondents may resent not having a neutral option, particularly a “don’t know” response category.⁵² As part of the pre-testing process, ask respondents about the scales used – would another set of responses be more appropriate? While research regarding the use of a “don’t know” option is inconclusive, your training evaluation survey should include a “don’t know” option if it’s plausible that people may simply not know, not have an opinion, or be offended by the forcing of a response.
- **Use 5-7 point rating scales.** Over time, there have been many discussions and disagreements in survey research focused on one central question: What works best with the Likert scale to give you the most accurate responses? Most measurement scholars agree that more than seven points on a scale are too much. Studies have shown that people are not able to place their point of view accurately on a scale greater than seven.⁵³ What is the perfect number? Studies are inconclusive, but the most commonly recommended scales are five, four or three point scales. Current thinking suggests that 5-7 point scales are adequate for the majority of surveys that use ordered responses or rating scales. However, conclusive evidence for the superiority of either odd or even numbered scales is currently unavailable. You should use whichever best suits your survey’s needs, but as mentioned above, 4 point or 6 point scales can be used to prevent artificial clustering around a midpoint. Regardless of which scale you use, be sure to pre-test your questions and response categories!
- **Use rankings only if respondents can see or easily remember all choices.** Rankings or rank-order scales are a type of ordinal measure in which choices are placed in a list and respondents are asked to order them from the highest to the lowest (or the other way around). The rank of training priorities among a list of training options, for example. In self-administered written surveys or online survey questionnaires, it’s important that the list of options be readily visible in their entirety and not be broken by a page. If lists of alternatives are too long, you may want to consider asking respondents to choose the top two or three and the bottom two or three.

⁵¹ Krosnick, J.A. (1999).

⁵² Babbie, E. (2009)

⁵³ Babbie, E. (2009); Krosnick, J.A. (1999)

SECTION 2.4

Online or Web-based Surveys

Web-based or internet surveys are becoming more common training evaluation tools. Survey instruments that respondents complete on the internet look like other self-administered questionnaires. The rules for writing questions for online surveys are almost exactly the same as those that apply to other self-administered questionnaires. They include making sure that you have a specific purpose for the question you ask, that you understand the needs of the survey's users, and that you have the resources available to conduct and complete all survey activities. Also, when you conduct any type of survey, you must respect the cultural and sociopolitical beliefs of respondents and their ability to understand and complete each question.

On the whole, the types of questions that are appropriate for online surveys are fairly similar to those used in other self-administered questionnaires. In fact, an important and ongoing area of survey research is whether online surveys are simply a technological advance in self-administered survey design or if they are actually a different type of survey altogether.

When Should you Consider an Online Survey?

Internet surveys may be preferable to written surveys in the following cases:

- The survey can be conducted with a convenience sample. In a convenience sample, the probability with which a respondent is selected into the sample may not be known – the respondents “self-select” into the survey.
- You have a list of e-mail addresses for the target population. The benefits in terms of cost and timeliness are greatest when the target population can be contacted to receive the survey via e-mail.
- The sample size is relatively large. Generally, web-based surveys have a larger initial start-up cost than written self-administered surveys. However, web-based surveys are more cost effective at the analysis stage with larger target populations, as web-based survey programs have automated data tracking, coding and analysis features.
- The survey contains questions of a particularly sensitive nature. There may be a bias toward socially acceptable answers (as opposed to more honest answers that may be less socially acceptable) in surveys in which the respondent has direct contact with the researcher.⁵⁴
- The survey contains a large number of important open-ended questions. Web-based surveys incur no coding or editing costs because responses are received electronically. There is also some evidence

⁵⁴ Babbie, E. (2001).

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

that respondents give longer answers to open-ended questions in electronic surveys than open-ended questions in written self-administered surveys.⁵⁵

- The survey includes graphics or contains interactive elements. A web-based survey easily allows the use of color and simple graphics to make your survey more attractive and professional-looking. Although online surveys rely on the same principles of question writing as do other self-administered surveys, the question and response formats you can use online are more varied. For example, questions in online surveys can offer responses in dropdown lists. In addition, while complicated skip patterns may appear confusing to respondents to paper surveys, surveys that require complicated skip patterns can be easily built into web-based surveys and you don't have to rely on the respondent's ability to follow your skip instructions.
- The survey requires some responses. There may be a reason, based on your evaluation objectives, to require a response to a question before an individual can move on to the next question. This is easily accomplished in an online survey.

Guidelines for Designing and Implementing a Web-based Survey

The recommendations for designing effective surveys (question construction, order, response categories, pre-testing, etc.) that we've already discussed apply to online or web-based surveys as well. However, to implement an effective online training evaluation survey, we also recommend the following:

- List only a few questions per screen -don't require the respondent to scroll down too much on a single page.
- Eliminate unnecessary questions -use your logic model and training objectives to delete any extraneous questions.
- Use graphics and color sparingly – too many graphics and too many colors can be distracting.
- Use matrix (table or grid) questions sparingly. Or, if a matrix is used, don't use too many lines in the matrix table - this may make your survey look unnecessarily difficult.
- Reduce response errors by restricting response choices. Make sure any critical questions to the analysis require responses before individuals can move on. Require responses for open-ended questions as well if they are critical to providing you with further clarification or explanation of responses.
- Be careful not to over-use the "force" choice option - use it only when a response is required for assessment of a specific learning objective.
- Ensure that respondents' privacy and their perception of privacy are protected.
- Provide some indication of survey progress (e.g., a bar graph that shows the percentage of the questions completed or the percentage of the survey remaining to complete).

⁵⁵ Heergweth (2005).

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

- Provide respondents with an estimate of how long it will take to complete the survey. Use pre-test information to make this estimate as accurate as possible.
- Allow respondents to interrupt and re-enter the survey.
- Inform respondents of the deadline for completing the survey (e.g., let them know when the survey will close).
- Remember to send a follow-up email to encourage respondents to complete and return the survey.
- Use automatic skip patterns (build these into the program rather than only in the instructions to respondents).
- Thoroughly pre-test the survey.
- Provide a means for respondents to report problems they are having with accessing or completing the survey.

SECTION 2.5

Interview and Focus Group Methods

Qualitative methods such as interviewing and focus groups are important training evaluation tools that not only can provide more detailed information about satisfaction with the training experience, but also can provide detailed information about learning acquisition and behavior and practice change. By listening to individuals, you can learn about important aspects of participants' experience, learning, and application of knowledge that you may not have known about before or can capture through quantitative means.

Focus groups and interviews are used in evaluation research methods to collect qualitative data. The utility of including interviews and focus group methods in your training evaluation depends on asking good questions. While interviews can vary in formality (from informal to structured interviews), interviews, just as surveys, need to carefully frame questions and ask all the respondents the same questions in the same words in the same order to provide reliable data for your evaluation.

Focus groups can be especially helpful to providing training organizers with more in-depth information about training experiences (as well as learning and application of new skills, policies and procedures in the field). The purpose of a focus group is to develop an understanding of the participants' experiences or views about an activity, a program, or a product or service. The purpose of a focus group is not to reach consensus, provide recommendations, or make decisions. Typically a focus group is composed of 8 to 10 people, who are selected because they have certain characteristics or experiences in common. Participants are asked to share ideas and perceptions about a particular area of interest in a relaxed, comfortable atmosphere. The discussion is carefully planned around a few open-ended questions. The focus group moderator plays an unobtrusive role, encouraging comments, both positive and negative, and being careful not to make judgmental comments. At the opening, the moderator ensures that participants understand that they are expected to express opinions and feelings and that alternative explanations are expected. Participants are assured that there are no right or wrong answers, only alternative points of view. The moderator promotes interaction and assures that the discussion remains on topic. The interview follows pre-determined questions arranged in a logical order.

In collecting data needed for training evaluations, focus groups have a number of advantages over surveys or individual interviews. Focus groups are a rich source of information because participants' comments build on the insights of other participants. A wider range of insight is provided because a comment by one person often triggers a chain of responses. Focus groups provide a quick and cost effective method of gaining a large amount of information.

Important Considerations when Implementing Interview and Focus Group Methods

Be Consistent and Neutral. In conducting interviews for *evaluation purposes*, it is important to be consistent and neutral. Just as you want to reduce bias in quantitative measurement instruments, you want to reduce bias when conducting qualitative interviews. Structured interviews (with pre-set questions, including agreed upon probes to encourage detail) facilitate treating all participants in the same manner during the interview. However, whether your interview is structured, semi-structured, or more informal, to reduce bias, all interviewers need to be consistent and neutral in their interviews of individuals.⁵⁶ If you do not treat all training participants the same way in an interview, your conclusions about participants' experiences may be undermined by any differences in how you treated the interviewees. For example, you may find that program participants who mentioned positive aspects of the program talked in more detail during interviews than participants who mentioned negative aspects. But your findings may be undermined by the fact that you may have, even inadvertently, encouraged participants to talk more when they mentioned something positive about the training program by using reinforcing statements such as "I think you've mentioned a really good point." On the other hand, you may have discouraged participants from expanding on their points when they mentioned something negative by using statements such as "I'm disappointed to hear that."

To be Consistent, interviewers should:

- Always read the instructions to each participant as stated.
- Ask every question as stated.
- Ask the questions in the order stated.
- Use standard phrases in response to a participant if he/she does not understand a question, or gives tangential or overly lengthy answers.

To be Neutral, interviewers should:

- Avoid agreeing or disagreeing with a participant.
- Avoid indicating that a participant's answer is 'right,' 'wrong,' 'good,' 'poor,' or 'interesting.'
- Avoid suggesting an answer or interpreting a question for a participant.
- Avoid giving your own opinion.
- Use standard phrases in response to a participant who is 'chatty' or asks for the interviewer's opinion.

⁵⁶ Patton (1987).

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

Ensure a Good Flow for Questioning. One challenge in reading all of the questions in the same way, is the need to use a smooth and conversational tone in all parts of the interview that are read to respondents (including instructions, probes and prompts). Even formal interviews should flow like a conversation. Avoid organizing interview items in a way that requires the interviewer to go back and forth in the interview form. Generally, the first items in your interview should establish the purpose for the interview and maintain the respondent's interest. As a result, any demographic questions that need to be obtained during the interview should not be among the first questions. When the questions flow logically from the introduction, respondents are drawn into the interview rather than being distracted and annoyed by questions they may consider irrelevant. A smooth start also sets the tone for the rest of the interview, establishing a "rapport effect" that builds trust and enhances the respondent's willingness to participate fully in the interview. Include transition statements that tell the respondent what topic the interviewer is going to address next. Transition statements give respondents a sense of movement through the interview and establish an overall coherence among the parts of the interview. Again, when designing your interview instrument, you must consider the possibility of question-order effects, or situations in which answers to certain questions may influence respondents, consciously or unconsciously, in their responding to later items.

Identify Information-Rich Participants. Think carefully about the characteristics of the people you should have in your interviews or focus groups. Who will give you the most in-depth information to supplement any quantitative data you have about the training? Are there specific groups that you need to hear more from? It sometimes helps to think of this as

Sample Interview transition statements:

First, I'd like to start by asking you some general questions about your training experience ...

Now, I'd like to learn more about specific workshops you attended ...

Next, I'd like to ask you for your recommendations for how the training program can be improved ...

We're almost finished with the interview; I just have a few questions about ...

This completes our interview. Thank you for taking the time to participate - do you have any comments you'd like to add?

The general rule of thumb for the number of focus groups in evaluation is to plan for 3-4 focus groups. Once you have conducted these, determine if you have reached "saturation." Saturation describes the situation when you have heard the range of ideas or issues and aren't getting new information. If you are still getting new information after 3-4 groups, you may need to consider additional groups.

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

identifying the “information-rich” respondents.⁵⁷ After determining what you need to discover through the interview or focus group process, ask yourself “Who has the greatest amount of insight on this topic?” Include those individuals in interview and focus group methods.

Steps in the use of Focus Group Interviews

1. *Developing a questioning route.* Questioning route is known as the order or sequence in which questions will be administered. The questioning route should be based on the objectives you have defined for the focus group. Brainstorming with colleagues and prospective users of the information can then help you to generate the questions.
2. *Recruiting the participants.* Participants for the focus group should be those individuals who are most likely to provide you with in-depth information on the target issue, or those individuals from whom you have not had input (i.e., an under-represented group). Participants should also be selected on the basis of their ability to discuss freely in a group, and their interest in the topic.
3. *Planning resources.* Develop a timetable for the focus group, including planning, implementation, and evaluation and reporting. Consider the resources both in staff, time and cost that will need to be marshaled for the focus group. Secure a comfortable facility for the focus group - one that includes sufficient space for flip-charting the discussion.
4. *Moderating interviews.* Ensure the moderator is someone that is comfortable and skilled at facilitating group discussion and is well-versed in the objectives of the focus group study. Moderators must keep the discussion on track. Moderation involves bringing the conversation back on target when irrelevant topics are introduced. This guidance has to be provided without reducing group enthusiasm and interest in the discussion. There are several personal attributes of a good moderator,⁵⁸ including: familiarity with group process either from previous experience in working with groups or through training in group dynamics; good listening skills; adequate background knowledge on the topic of discussion; well-developed written and oral communication skills; and a sense of humor.
5. *Data analysis and reporting.* Data analysis and reporting can follow an interpretative summary format, whereby the data are not only described but also interpreted.⁵⁹ This analysis produces an interview summary including key incidents, strong statements, and frequently occurring responses. Next, the key incidents, strong statements, and frequent responses are classified by question, coded, and grouped. The coding and grouping helps to identify the general themes in the responses.⁶⁰ The identified themes are then compared across interviews (or series of focus groups) in order to develop a general picture on a question by question basis and to draw conclusions related to the objectives of the focus group study.

⁵⁷ Patton (1990).

⁵⁸ Krueger & Casey (2000)

⁵⁹ *Ibid.*

⁶⁰ Miles & Huberman (1994)

SECTION 2.6

Survey Sampling

Ideally, surveys of training participants about their satisfaction with and reactions to the training would include the entire population of training attendees. In this situation, a sample would not be needed. However, it may be the case that you have to “sample” the training population for survey measurement - and also to sample the training participants to receive additional follow-up measures that cannot realistically be applied to all of the training attendees (due to cost, time constraints, etc.). All samples are subject to error (called “sampling error”), though this error can be minimized by including larger samples and samples of individuals that are homogenous (similar).

Convenience and Probability Samples

There are two main types of survey samples: convenience samples and probability samples. Convenience samples arise from uncontrolled instrument distribution (not everyone who could receive a survey receives the survey) or self-selection (i.e., individuals self-select to take and complete the survey). Probability samples, also called random samples, are samples in which the probability with which an individual was selected into the sample can be determined. Probability samples are purposely random - each person in the survey population has an equal chance of being included or excluded from the survey sample. Probability samples can be classified into three types: those taken from closed populations (such as the list of individuals who attended a training), general populations, and pre-recruited panels. Convenience samples are often less costly to generate than probability samples, but the statistical inference needed to generate conclusions from those samples becomes problematic.

While the nuances of survey sampling are beyond the scope of this *Guide*, the following matrix provides definitions and examples of when different approaches might be used.

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

Table 2-2 Probability Sampling Designs	
Simple Random Sampling (SRS)	Each person in the survey population is assigned a random number, which is then randomly selected for inclusion in the sample.
Systematic Sampling	The survey population is systematically numbered 1-X, and each Nth person is selected. For example, each 10 th person is selected for inclusion in the sample. This differs little from SRS, but it is more accurate in many instances. One danger to consider is periodicity: if the survey population is coded cyclically, (e.g. - in an order), the sample may be biased. A second danger to be aware of is implicit stratification: if the persons in the survey population are arranged in a certain pattern (e.g. - alphabetically by last name).
Stratified Sampling	Instead of drawing the sample from the entire survey population, the survey population is broken down into categories that are then sampled appropriately (though not necessarily equally). For example, a population might be broken down by age before being sampled. Stratified sampling produces a more representative sample from the survey population, markedly reducing sampling error.
Multistage Cluster Sampling	Cluster sampling is best used when the survey population is too large to list individually for sampling (e.g. – all judges in the United States). It involves listing possible data sources (or clusters) and either sampling or stratifying those clusters until a representative, usable sample is defined. For example, you might list all states in the United States and then sample or stratify the states into one cluster; then you might list all counties in each selected state and sample or stratify the counties into another cluster; and so forth, until you arrive at a reasonable and representative sample.
Probability Proportionate to Size (PPS)	In some instances of Multistage Cluster Sampling, very large populations stand a chance of being overrepresented, and very small populations stand a similar chance of being underrepresented. In such cases, combining several small populations into one larger cluster is common practice, as is pre-sampling larger populations to account for their overrepresentation.
Non-probability Sampling Designs	
Purposive Sampling	Purposive sampling involves the selection of a sample based on the researcher's knowledge of the larger community or sampling frame. Though this provides very well-articulated results, they often fall victim to biases of the researcher.
Quota Sampling	For quota sampling, the researcher must have considerable demographic data about the sample population. The sample population is broken down into demographically-delineated groups which are then appropriately weighted according to their portion of the total population, resulting in a reasonable representation of the sample population.

TOOLS AND RESOURCES – CHAPTER TWO

- Checklist to Determine if a Formal Instructional Design Approach is Needed
- Curriculum Design Worksheet
- Framework to Design a Training Plan
- Overview of Data Gathering Methods for Needs Assessment
- Personal Learning Styles Inventory

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

CHAPTER TWO: References and Resources

- Babbie, E. (2009). *The practice of social research*. Wadsworth.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for transfer. *Psychological Bulletin*, 128, 612–637
- Cree, V., & Macaulay, (2000). *Transfer of learning in professional and vocational education*. Routledge.
- Dillman, D.A. (1978). *Mail and telephone surveys: The total design method*. New York: Wiley.
- Dillman, D.A. (2000). *Mail and internet surveys, 2nd Edition*. New York: Wiley.
- Fowler, F.J. (1988). *Survey research methods*. Newbury Park: Sage.
- Fowler, F.J. (2001). *Survey research methods, 3rd Edition*. Thousand Oaks: Sage.
- Heerwegh, D. (2005). Personal salutations and email survey response rates. *Public Opinion Quarterly*, Vol. 69(4), 588-598.
- Herberlien, T.A. & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature, *American Sociological Review*, Vol. 43 (4), 447-462.
- Krosnick, J.A. (1999). Survey research. *Annual Review of Psychology*, 1999, Vol. 50, 537-67.
- Krueger, R.A. & Casey, M.A. (2000). *Focus groups: A practical guide for applied research*. Newbury Park: Sage.
- Martin, C. L. (1994). The impact of topic interest on mail survey response behavior. *Journal of the Market Research Society*, 36 (4), 327-337
- Miles, M. B., & Huberman, M. A. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage Publications.
- Patton, M.Q. (1987). Depth interviewing. In *How to use qualitative methods in evaluation* (pp. 108-143). Newbury Park: Sage Publications.
- Salant, P. & Dillman, D.A. (1994). *How to conduct your own surveys*. New York: Wiley, John & Sons, Inc.
- Schunk, D. (2004). *Learning theories: An educational perspective* (4th ed.). Upper Saddle River, NJ, USA: Pearson
- Schwarz, N. & Sudman, S. (1996). *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. Jossey-Bass.
- Sheehan, K. B., & McMillan, S. J. (1999). Response variation in e-mail surveys: An exploration. *Journal of Advertising Research*, 39 (4), 45-54.
- Steele, T. J., W. L. Schwendig and J. A. Kilpatrick. "Duplicate Responses to Multiple Survey Mailings: A Problem?" *Journal of Advertising Research*, March/April, 26-34.
- Sudman, S. & Bradburn (1982). *Asking questions: A practical guide to questionnaire construction*. Jossey-Bass.
- Thagard, P. (2005). *Mind: Introduction to Cognitive Science*. Cambridge, Massachusetts: The MIT Press.

CHAPTER TWO: TRAINING SATISFACTION AND REACTION MEASUREMENT

Watt, J. H. (1999). Internet systems for evaluation research. In G. Gay & T. Bennington (eds), *Information technologies in evaluation: social, moral, epistemological and practical implications* (pp. 23-44). San Francisco: Josey-Bass, no. 84.

Yammarino, F.J., S. Skinner and T. L. Childers. "Understanding Mail Survey Response Behavior." *Public Opinion Quarterly*, 55 (1991): 613-639.

CHAPTER THREE

Measuring Learning Acquisition

The second level of evaluation involves measuring learning. Evaluating learning is more complicated than evaluating satisfaction because it requires more in-depth measurement on behalf of evaluators including comparison of learning after training against a baseline or pre-training measure. The use of quasi-experimental designs (pre- and post-test comparisons) is essential in true learning acquisition assessment. If possible, for example, participants might take a knowledge test or skill assessment before the training (pre-test) and after training (post-test) to determine the amount of learning that has occurred. In addition, to determine whether learning has transferred beyond the training to the workplace, additional measurement is needed using tools such as observation and peer or supervisor feedback.

Essentially, assessing at this level moves the evaluation beyond learner satisfaction and attempts to assess the extent students have advanced in skills, knowledge, or attitude. Methods range from formal to informal testing to team assessment and self-assessment. The materials presented in previous Sections (e.g., instructional design, learning styles, principles of adult learning, the development of measureable learning objectives, and the survey process) are foundational to learning acquisition measurement and should be revisited when working through the concepts in this Section. The materials presented in Chapters Four and Five of this Guide are also foundational to the measurement of learning (i.e., experimental and quasi-experimental methods) and readers should apply the concepts presented in those chapters to their understanding of learning acquisition assessment.

This Chapter considers what we mean by “learning,” the types of learning that can occur during training programs, the importance of your instructional design model to learning acquisition, and the methodologies evaluators can use to properly measure learning acquisition. Quasi-experimental design (pre- and post-test measures as well as the use of control groups) is introduced in this Chapter and will be further discussed in the Chapters on behavior change and impact or results measurement.

SECTION 3.1

Learning Acquisition Evaluation Basics

The Domain of Measurement - What is Learning?

To most of us, learning refers to knowing something. But psychologists do not agree with this simplistic view about learning. Although there is no single definition of learning, a generally accepted definition of learning is any relatively permanent change in behavior that occurs as a result of experience.⁶¹ This means that an external observer has to recognize that learning has taken place. Belkin and Gray (1977) define learning as a change in the individual as a result of some intervention. It may be viewed as an outcome or as a process. Rogers (2003) views learning as task-conscious or acquisition learning (e.g., the learning involved in parenting). Formalized learning arises from the process of facilitating learning. It is educative rather than accumulation of experience. Formalized learning makes learning more conscious in order to enhance it. Smith (1982) views learning as a product (the acquisition of a particular set of knowledge), process (how learners seek to meet needs and reach goals), and a function (how learners are motivated, what brings about change).

Training means a set of well-defined actions undertaken to achieve predetermined goals. Training is goal-oriented. Learning in the context of training should be associated with post-learning application.

Training is an instructor-led, content-based intervention, leading to desired changes in behavior.⁶² In training, learning is viewed as an intervening variable to cause behavioral change, which is a dependent variable and the training experience works as the independent variable. Two processes, or stages, of learning in the context of training are evident – the process of acquiring skills, knowledge and concepts, and the process of putting these into actions. This differentiates training from education. In fact, training means a set of well-defined actions undertaken to achieve the predetermined goal, while in educating neither the objective is necessarily given nor is the means of getting it distinct.⁶³ Training is goal-oriented and, unlike education, each action is pre-scheduled. Learning in the context of training, therefore, is well connected with post-learning application, otherwise known as the transfer of learning. Learning is an integral part of training.

⁶¹ Robbin (1998)

⁶² CIPD (2005)

⁶³ Skinner (1968)

Three Types of Learning to Measure

What exactly trainees learn from a training program varies according to the purpose of the training. Trainings are conducted to introduce new information, but some go further, and attempt to introduce trainees to a new set of skills or change or shape trainees' attitudes toward an important occupational issue. Yet in spite of the specific goal of training, each of these three types of learning may occur to some degree.⁶⁴

Knowledge Acquisition – All trainings, at some level, provide new information to trainees that will enable them to enhance their job performance. An example would be a training program designed to inform legal practitioners of new laws. In these types of trainings, the goal for training coordinators is to ensure that the appropriate information is retained by trainees so they have the opportunity to apply it to their job. If training organizers believe retention of this material is important, as they usually do, it is important to measure the knowledge trainees acquire as a result of the training program. Evaluating knowledge acquisition requires evaluators to apply pre- and post-test methodology, measuring participants' knowledge prior to and after receiving the training.

Skill Acquisition – The goal of some training is to teach trainees new skills that will enable them to enhance their job performance. An example would be a training that discusses how attorneys should interact with their clients. Here introducing the trainee to a set of skills is the primary goal of the training. Acquiring new skills does not only include the incorporation of new behaviors, but can also include the elimination of disruptive behaviors that inhibit job performance. Therefore training organizers are often interested in measuring whether trainees have acquired new skills as a result of the training. Ideally, evaluating skill acquisition requires evaluators to obtain a baseline of skill-level before individuals are trained and then re-testing individuals on that skill set after the training has occurred. The extent to which skills are present or absent (both before and after training) can also be corroborated by external sources such as supervisors, peers or mentors.

Attitude Change – Some training is directed at providing trainees with a new perspective. Here trainees are asked to change their perspective regarding some key aspect of their occupation. Some training programs may try to introduce trainees to a new topic. An example of current training efforts on attitude shaping are trainings designed to introduce the importance of dental health care for dependent children. This is a topic not readily thought of, and creating buy in among practitioners is important in establishing a positive attitude toward overcoming the issue. Other trainings are directed at changing harmful attitudes, such as overcoming interpersonal issues that are impeding job performance. An example is current training efforts on working with lesbian, gay, bisexual, transgender, questioning, intersex, and asexual (LGBTQIA) youth in the child welfare system. While changing trainees' attitudes is not always the primary objective of a training program, attitude change is required at some level to facilitate knowledge acquisition, and is important to consider in any evaluation. To determine attitude change, evaluators would ideally obtain measures of trainees' attitudes

⁶⁴ See for example Chowdray (2006); Rogers (2003); Robbins (1998); Smith (1982); and CIPD (2005).

CHAPTER THREE: MEASURING LEARNING ACQUISITION

on an issue or topic before the training and compare those attitudes to measures taken after the training has concluded. In addition, evaluators may seek additional evidence of attitude change via other sources of information (e.g., observations in the field, whether practices are implemented or changed, etc.).

The purpose of training programs is to expose trainees to information that will enhance their ability to do their job. While training may focus on instilling information primarily through one of these types of learning, each of the three types of learning occurs at some level in every training program. Training organizers should examine their training objectives carefully to determine if their learning objectives relate to knowledge acquisition, skill acquisition, or attitude change or some combination of these types of learning.

Measuring learning acquisition in the training evaluation context is essentially providing an assessment of the success of your instructional design – was your instructional design effective at facilitating and achieving knowledge, skill or attitude gains?

- Learning is facilitated when learners are engaged in solving real-world problems.
- Learning is facilitated when existing knowledge is activated as a foundation for new knowledge.
- Learning is facilitated when new knowledge is demonstrated to the learner.
- Learning is facilitated when new knowledge is applied by the learner
- Learning is facilitated when new knowledge is integrated into the learner's world.

SECTION 3.2

Methods for Measuring Learning

How do we know that Learning has Happened?

When learning is evaluated, there are many questions to answer. Central of course is how learners experience the learning process and what they actually learn (the outcome of learning), their knowledge, attitudes, skills, and aspirations, and their behavior change. An evaluation also can focus on the educator as learner, and the content and resources that are used by the trainer and what he or she learns about the training process. Because learning is always a social phenomenon, an evaluation can focus on the social environment, organizational context, and the relevance of language, culture, and sometimes public policy to learning. These underlying cultural assumptions often explain resistance to learning, as well as the way learning either reproduces existing racial, gender, and economic power relationships or challenges these relationships. Not all evaluations include all of these questions. Training organizers tend to focus on questions that serve their own perspectives. Learners may also be interested in questions that serve their own perspectives.

The Perspective of Learners

When training participants consider evaluation of their own learning, they may ask themselves a broad range of questions. For example, have they, as learners:

- Gained knowledge or problem-solving skills that are useful to them?
- Increased their hopes and aspirations regarding systems change?
- Learned how to gain better access to more knowledge?
- Changed their assumptions, habits, pre-conceived opinions or priorities?
- Gained confidence in taking leadership to make change?
- Increased their commitment to experiment or take direct action?

The Perspective of Training Organizers

When training organizers consider the evaluation of learning, they usually want to know how the learners perceive the process of learning, especially how they, as trainers, have been helpful to the learning process. Trainers ask themselves - and ask the learners to indicate - whether they have:

- Met expectations and objectives?
- Introduced a variety of useful methods, strategies and materials?
- Encouraged the use of examples to illustrate concepts or practice?
- Given step-by-step instructions for the application of methods, strategies, concepts to the field?
- Summarized the material presented?
- Related theory to practice?

CHAPTER THREE: MEASURING LEARNING ACQUISITION

- Showed concern about the learners as professionals and as human beings?
- Promoted discussion and learner interaction?
- Encouraged silent learners to participate?
- Used vocabulary that is readily understood?
- Respected racial, ethnic and gender differences and any unique contributions to learning?
- Appreciated any barriers or challenges to learning?
- Helped learners to reflect critically on how they learn?
- Appreciated the local knowledge and expertise of learners and made use of it in training?
- Helped learners learn from each other during learning activities?

The purpose of listing all of these questions from both the learners' and trainers' perspectives is not that one should ask them all, but to encourage training organizers to think about which ones are essential items for a specific evaluation effort. Re-visit your training logic model to prioritize learning measurement.

Methods that Gather Evidence of Learning

Ensuring participants learn the material presented to them is important not only because it is typically the goal of a training program, but also because it is a pre-requisite for behavior change. People cannot change their behavior if they did not learn something. Evaluations should include measures of learning whenever possible, as it is a vital component in ensuring the utility of a training program.

Documentation of Prior Knowledge – A basic principle of adult education is that learning should begin with prior knowledge. How can we appreciate what has been learned if we don't know what participants already know? One method trainers can use to gather existing knowledge is to ask training participants to document what they know (in written form via a skill test or questionnaire or orally via an interview). This is essential, since evaluation of new learning should be compared with what training participants already know. Also, training participants' acknowledgment of the limitations of their existing knowledge can motivate them for the training experience.

The most general way to measure learning is to ask participants how knowledgeable they are on the topic of interest. If training is focused on teaching trainees new Indian Child Welfare Act protocols, you could ask them directly how knowledgeable they are on the topic before the training on a rating scale that ranges from “no

Basic Example:

Before Training: How knowledgeable are you of the provisions of the Indian Child Welfare Act?

Not Knowledgeable 1 2 3 4 5 Very Knowledgeable

After Training: How knowledgeable are you of the provisions of the Indian Child Welfare Act?

Not Knowledgeable 1 2 3 4 5 Very Knowledgeable

CHAPTER THREE: MEASURING LEARNING ACQUISITION

knowledge” to “very knowledgeable.” Ask trainees to indicate their knowledge level before and after the training and measure the increase.

Training organizers can use rating scales and checklists to assess the knowledge of learners. Learners can also use these to check the performance of trainers - how in tune are trainers with learners’ educational needs and gaps in knowledge? Rating scale and learning checklists can be administered in individual, group or field-settings or online. Not only are they valuable measures for training organizers, but learners can use them to judge their own performance, current knowledge, or educational expectations.

Measure the Specific Material Taught – While it is possible to rely on trainees’ perspectives of knowledge acquisition (self-report), it is not the best measure. This is because human estimates of knowledge are extremely fallible. In addition, it does not provide a detailed assessment of the specific learning that took place. The trainees said they learned something, but what exactly did they learn? Determining what specific knowledge was acquired requires evaluators to test trainees regarding the information they learned. This requires evaluators to design items that resemble a quiz, asking a series of questions regarding the material that was just presented to them in the training. This test will need to be administered to the trainees before and after the training, to meet the pre- and post- test requirements to measure learning. Evaluators can give the same test twice, as long as they do not reveal the answers in the first test. Although more time consuming, it is optimal to make up different tests with similar questions regarding different aspects of the material, therefore eliminating the possibility that testing effects will influence evaluators’ measure of learning. In addition, training organizers should consider post-tests of learning that involve scenarios, role plays, or problem-solving. These measures not only assess understanding and knowledge acquisition but they also provide a means to determine if trainees have learned how to apply that knowledge to a problem or the real-world setting.

Basic Example [Ask both pre and post training]

According to the Indian Child Welfare Act, the tribe has jurisdiction over a child removed from their home?

- ☐ True
- ☐ False

Please mark which of the following parties has the right to an attorney according to the Indian Child Welfare Act?

- ☐ The Child
- ☐ Non-Abused Siblings
- ☐ The Child’s Native American Parents
- ☐ The Child’s Non Native American Parent
- ☐ The Tribal Custodian
- ☐ Other [please specify]: _____

In spite of the negative attitudes associated with tests and quizzes, they can be useful for diagnosing learner proficiencies, documenting prior knowledge, projecting learning achievements, and understanding learner attitudes. Repeating a quiz given at the outset of the training at the end of a learning event not only documents change, but also serves to demonstrate to learners themselves what they have learned.

CHAPTER THREE: MEASURING LEARNING ACQUISITION

Retention: Evaluating retention involves the addition of follow up measures some time after the training. To ensure trainees retained the information they learned, training organizers must send a follow-up evaluation a few months after the initial training program. Again, this measure may resemble a quiz, inquiring about trainees' knowledge level on several aspects of the original material presented to them.

Creative Methods to Assess Learning

Case Studies – The use of case studies enables participants to apply what they learn to a real life scenario. Working in teams, participants use a case study to apply what they learned to solve problems relevant to the skill sets, strategies, or concepts covered in the training.

Observation – By observing how the participants apply their new skills, a facilitator can understand how well the participants will be able to apply what they learned back on the job. For example, if the program focuses on improving presentation skills, the facilitator may have each participant develop a presentation and record them giving that presentation to the class. By videotaping the presentation, feedback can more effectively be provided to participants and they will have something they can walk away with to use later to improve upon their skills. In this example, you might use a “pre-test” (where you have the participant do a presentation before the training program begins and videotape him/her giving that presentation) and then a “post-test” after the training program ends (where you have the participant do a presentation after he/she has learned best practice skills for making presentations and videotape him/her giving that presentation) to show the improvement based on the training program.

Role Plays – Role plays are another way to have participants practice, in a safe environment, the skills they are learning in the class. Role plays work well with a variety of topics such as hearing practice, conflict management, mediation and leadership. Depending on the situation, you may allow the participants to develop their own situation to role play (based on an upcoming issue they must address or an area where they feel they have the most difficulty) or provide a scripted role play situation for the participants. Redacted court case files can be an excellent source to develop role plays to demonstrate learning with respect to hearing best practices.

Simulations – A computer-based simulation simulates a real life situation with all of the challenges and difficulties that are common. For example, a project management simulation will enable participants to work in a team environment to work through the challenges of strategic planning, project design and implementation, and project completion. The team may encounter challenges that it would find in any real project – such as loss of team members to other projects, reduction in their budget, or a shorter timeline to meet benchmarks.

Feedback Committees – When training takes place, participants can elect a “feedback committee” to provide learning observations to training organizers - the extent to which learning is both facilitated during the training,

CHAPTER THREE: MEASURING LEARNING ACQUISITION

and in the committee's opinion, whether learning is occurring. Members of the feedback committee, for example, can observe the training and use a structured observation form to provide an assessment of the degree to which learning is occurring. The feedback committee members may also interview supervisors or peers after the training to determine if learning occurred. They may also return after a number of months to re-interview supervisors or peers to determine if that learning has been retained. The feedback committee should also be open to any complaints, ranging from relevance of content, adequacy of facilities, or effectiveness of leadership, to the involvement of learners in discussion or activities, that participants may have about the training event,. The committee can bring items to the attention of the training organizers through written or verbal forms.

Group Discussion Assessment – This method can be incorporated into the training program itself. Training participants are broken up into small groups and then meet to discuss what they have learned (or have not learned). Small groups then complete a feedback form to return to training organizers. Small groups may also be asked to report back to the entire training group.

Peer Review Panels – Training participants can become involved in evaluating one another's learning through peer review panels. Panels can be taught to use standard forms and rating scales. Their evaluative judgments can be made with or without identification of reviewers. When peer review panels are used, it is important to establish a positive climate of constructive criticism.

End-of-Event Analysis – This is an assessment of learning at the conclusion of the training event and can be done in several ways. The most frequently used is an evaluation form that is administered at the end of workshop sessions, asking participants to gauge their level of learning. Another way is to have these forms distributed, collected, and summarized by a feedback committee. The feedback committee can then report these findings to the training group as a whole at the conclusion of the training and could facilitate a discussion on the overall strengths of the workshop or training event.

Testimonials and Stories – Testimonials and stories, in which training participants are asked to describe their learning experience through narrative means, can provide subjective records of individuals' educational experiences and the learning that takes place from the perspective of the learners. They can qualitatively describe the nature and process of educational change. These stories also can be easily understood by others outside the training program as illustrations of types of outcomes and can lead to ideas for future programming. Disadvantages of this method include social desirability bias (individuals may be pre-disposed to present themselves in the most favorable light through their stories), non-generalizability beyond the person giving the testimony, and difficulty in determining what happened as a result of the training program versus other influences on the person. Stories can be either written by the learners or created as a result of an interview or focus group.

CHAPTER THREE: MEASURING LEARNING ACQUISITION

360 Degree Feedback – In the 360 Degree Feedback technique, feedback about training participants' understanding and application of a new skill or knowledge is obtained from multiple sources such as the individual him or herself, his or her supervisor or manager, and his or her peers. Standard feedback forms can be prepared, and the data obtained can be used to measure changes in learning that occur after the training. Resulting feedback, for example, will indicate whether the individual trained exhibits specific skills or knowledge, whether change has occurred in skill set or understanding after the training event, and to what extent the training was successful in helping participants transfer learning to their actual workplace. An important consideration is to allow enough time between the end of the training and the 360 degree assessment so that participants have had sufficient time to put their learned skills into practice.

Round Robin Technique – In a training program where there are multiple workshops over time (e.g., participants attend one workshop and then return to complete a series of related workshops on the same topic at a later date), training organizers can use a round robin technique to facilitate the transfer of learning. In the round robin technique, trainees begin the second workshop in a series with a report on what they have learned and applied since the last workshop. Participants also talk about the success and problems they encountered in applying new skills or knowledge. This process is repeated for all of the workshops in the series - participants return to report in on what they have learned and been able to apply in practice. One advantage of this technique is that it supplies training organizers with information about obstacles to learning and application of knowledge in time to make curriculum adjustments to help trainees to overcome those obstacles. In addition, because trainees know that they have to return to future training workshops with a report of how they applied what they learned, they may be more likely to actually try out the newly acquired skills and knowledge in the workplace or field-setting.

SECTION 3.3

Guidelines for Accurately Measuring Learning

Measuring satisfaction only requires data to be collected at one point, sometime after the completion of the training. Measuring learning requires meticulous measurement and careful methodology. We have outlined four guidelines below that are required for evaluators to properly measuring learning:

Pre- and Post-tests – Self-report of learning is common in training evaluation feedback forms (e.g., at the conclusion of training participants are asked to report what they have learned or the degree to which they have learned specific concepts or tasks). But the evaluation of learning should not stop there. Evaluating learning involves two measurements, one before the training and one after the training. Trainees bring different levels of experience and knowledge to their job. Therefore it is important to measure trainees' knowledge level prior to the training. Measuring trainees' prior knowledge level not only allows organizers to calculate learning, but it also allows training organizers to tailor curricula to better meet the needs of the trainees.

Measure Each Type of Learning – It is important to measure each of the three aspects of learning at some level. Even if the primary goal of a training involves only one type of learning, with each of the three styles of learning occurring to some degree, it is important to incorporate some measures from all three. For example, if you are trying to teach dependency practitioners about ICWA statutes, there should be questions around knowledge acquisition specifically on ICWA guidelines. But there should also be items investigating skill acquisition and attitude change in order to understand trainees' perceptions regarding the depth of the training, and whether trainees plan to use the information they learned to change their practice (or to encourage others to change their practice).

Aim for a High Response Rate – If possible, try to ensure you collect responses from each trainee. The less people learn, the less motivated they will be to participate in the evaluation. Consider incentives to completing evaluations, including whether or not you should make evaluations mandatory as part of training attendance. Consider offering additional technical assistance materials for completing evaluations or providing opportunities for participants to serve as faculty for future trainings.

Use of Control or Comparison Groups – Employing pre- and post- test measures provides a solid estimate of learning acquisition. However, the optimal way to ensure that the training is resulting in learning is to conduct an experiment, employing the use of a control or comparison group. If possible, use a control group of co-workers/ peers or colleagues who did not receive the training. Your control or comparison group should be as similar as possible to your “experimental” or training group - with the only difference being the training intervention received by your training group (e.g., same level of experience, same jurisdiction, same supervisory structure, etc.). Ideally, a random assignment procedure of individuals to attend the training, or not to attend the training, would also be employed. Provide the control or comparison group and the training group

with the same measures and compare their knowledge. The use of control and comparison groups is covered in more detail in Sections Five and Six of this *Guide*.

Recommendations for Evaluating Learning Transfer

Since most dependency court training programs provide training to transfer learning back to the workplace or field-setting, the following recommendations apply for evaluating the efficacy of that learning transfer:

Assess Participant Satisfaction or Reaction

As mentioned in other Sections of this *Guide*, most training evaluations ask for participants to report how they felt about the training experience. While some research has indicated that how people feel at the end of training has little relationship with what they do back in the workplace (Alliger et al. 1997), it is still important to ask about participants' feelings. People tend not to learn very well if there are too many distractions (e.g., a noisy training venue), or if the training materials are presented badly, or the trainer is incompetent or ill-prepared. It is important to make sure that these issues do not get in the way of learning so they should be evaluated.

- To measure learning transfer, always include questions about the venue, the materials, the curricula and the trainer(s).

Assess the Utility of the Training for Participants

Alliger et al., (1997) demonstrated that how useful people say a training was for them is directly related to whether they will transfer their learning back to the workplace or field-setting. One of the simplest ways to evaluate whether or not a particular training program has been successful at increasing the probability of changing behavior back at work, is to ask participants, a scale of 1 to 10 "how useful has this training been to your current job?" A score of 1 would mean "no use at all," while a score of 10 would be "invaluable."

- To measure learning transfer, always ask participants a simple question about how useful the training has been to them in their current job or role.

Build Opportunities to Rehearse New Skills

Research has shown that there is a gain in work performance that can be directly attributed to whether or not an individual believes he or she can do a particular task or apply a skill in a given situation (e.g., Stajkovic & Luthans, 1998). This feeling of self-efficacy appears to be a major predictor of whether or not an individual will transfer learning from training back to the workplace or field-setting. In developing your instructional design, therefore, it's helpful to build in a number of opportunities for trainees to rehearse or practice the new skills you are teaching. For example, have trainees work through a role play of a real life situation they will face when they return to the workplace and have them apply the new skill or strategy. By rehearsing their new skills during the training program with a real situation in mind, feelings of self-efficacy are more likely to develop - and participants may be more likely to apply their new skills in the workplace.

CHAPTER THREE: MEASURING LEARNING ACQUISITION

- To encourage learning transfer, build in role plays and other strategies that provide training participants with opportunities to practice new skill sets.

Build Support for Learning Transfer in the Organization

Whether or not a trainee will transfer his or her learning back to the workplace or field-setting may be determined by the organizational culture the trainee is returning to after the training. The transfer of learning to the workplace can depend upon: the extent to which supervisors and co-workers encourage and set goals for trainees to use new skills or knowledge; whether there are features of trainees' jobs that serve as reminders to use the newly acquired skill or knowledge, and whether the organization offers implicit or explicit rewards (e.g., more responsibilities, greater credentialing, etc) to trainees to use the skills or knowledge acquired in the training. The organization must also provide opportunities for trainees to actually apply new skills and knowledge. Trainees may leave a training with the belief that they can do something, but without the opportunity to try it back in the workplace, their feelings of self-efficacy may dissipate, reducing the possibility that they will even attempt the new skill.

To facilitate the transfer of learning:

- Try to involve managers in the training design to address any obstacles to using new skills and ways overcome those obstacles.
- Make sure trainees have an opportunity to use new skills within a short time after returning to the workplace or field-setting. Work with leaders, managers, and supervisors to define and encourage those opportunities.
- Recruit leadership to support the application of the new skill or knowledge (e.g., Chief or Presiding Judge, Child Welfare Agency Administrator, etc.).

Evaluate Performance

The ultimate test of whether learning has been transferred is whether trainees are able to demonstrate their new skills or knowledge in the workplace or field-setting. To evaluate this transfer, you may use a 360 degree feedback technique to see whether the trainee, his or her manager or supervisor, and peers have actually noticed any changes in their practice. Do those "feedback" individuals identify the trainee as applying the same skill sets in practice that the individual was trained on? Have they noticed a change? Ideally, a pre-training measure as well as a post-training measure would be taken and the two compared - but, if that isn't possible, a single post-training questionnaire using a rating scale asking explicitly about application of skills and knowledge and change over time (e.g., does the trainee do more or less of X) may provide useful data about the transfer of learning.

- To evaluate the transfer of learning, consider using feedback from the individual as well as from his or her supervisors and peers.

CHAPTER THREE: MEASURING LEARNING ACQUISITION

- To evaluate the transfer of learning, consider using actual job performance evaluation data as part of your evaluation method.

TOOLS AND RESOURCES – CHAPTER THREE

- Sample Pre-Training Skill/Knowledge Assessment Form
- The Round Robin Interaction Method for Measured Learning
- Sample Questions to Assess Self-Report of Learning and Knowledge Acquisition
- Sample Post-Training Skill-Knowledge Assessment Form
- Sample Protocol for Testing and Reporting Learning Acquisition
- Sample Questions and Formats for Assessing Participants

CHAPTER THREE: MEASURING LEARNING ACQUISITION

CHAPTER THREE: References and Resources

- Alliger, G.M., Tannenbaum, S.I., Bennett, W., Traver, H. & Shotland, A. (1997). A meta-analysis of the relations among training criteria, *Personnel Psychology*, Vol. 50 (2), 341-358.
- Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Review*, 26(3), 201-213.
- Bass, B. M., & Vaughan, J. A. (1966). *Training in industry: The management of learning*. Belmont, CA: Wadsworth Publishing.
- Belkin, G.S., & Gray, G.L. (1977). *Educational psychology: An introduction*. W.C. Browne.
- Bramley, P. (1996). *Evaluating training effectiveness*. Maidenhead: McGraw-Hill.
- Bernthal, P. R. (1995). Evaluation that goes the distance. *Training and Development Journal*, 49(9), 41-45.
- Chowdary, M.S. (2006). Human behavior in the context of training: An overview of the role of learning theories as applied to training and development. *Journal of Knowledge Management Practice*, Vol. 7 (2).
- CIPD (2005) *Learning Styles, Fact sheet*, Chartered Institute of Personnel and Development, London.
- Dixon, N. M. (1996). New routes to evaluation. *Training and Development*, 50(5), 82-86.
- Gordon, J. (August, 1991). Measuring the “goodness” of training. *Training*, 28(8), 19-25.
- Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of the American Society of Training Directors*, 13, 3-26.
- Phillips, J. J. (1991). *Handbook of training evaluation and measurement methods*. (2nd ed.). Houston, TX: Gulf.
- Robbin, S. (1998) *Organization Behavior*. NJ: Prentice Hall.
- Rogers, A. (2003) *What is the difference? A new critique of adult learning and teaching*, Leicester: NIACE.
- Rossi, P.H., Freeman, H. E., & Wright, S. R. (1979). *Evaluation: A systematic approach*. Beverly Hills, CA: Sage.
- Ross, S. M., & Morrison, G. R. (1997). Measurement and evaluation approaches in instructional design: Historical roots and current perspectives. In R. D. Tennyson, F. Scott, N. M. Seel, & S. Dijkstra (Eds.), *Instructional design: Theory, research and models*. (Vol.1) (pp.327-351). Hillsdale, NJ: Lawrence Erlbaum.
- Skinner, B.F. (1968) *The technology of teaching*. New York: Appleton-Century-Crofts.
- Smith, R.M. (1982). *Learning how to learn: Applied theory for adults*. New York: Cambridge.
- This, L.E. and Lippitt, G.L. (1966) “Learning theories and training development,” *Training and Development Journal*, Volume 20, April-May, pp-2-11
- Worthen, B. R., & Sanders, J. R. (1987). *Educational evaluation*. New York: Longman.

CHAPTER FOUR

The Assessment of Behavior and Practice Change

An important goal of training is to ensure that behavior or practice changes occur as a result of the training – that training participants apply what they learned at the training. While an assessment of learning acquisition evaluates what was learned during the training, an assessment of behavior and practice change seeks to answer the question - *How will these newly acquired skills and knowledge impact performance?* This type of evaluation is more difficult and complicated to measure than evaluating satisfaction or learning.

This measures the transfer that has occurred in learners' behavior due to the training program, which attempts to answer the question - *Are the newly acquired skills, knowledge, or attitudes being used in the everyday environment of the learner resulting in a change in behavior or practice?* For many trainers this represents an ideal assessment of a training program's effectiveness. However, measuring at this level is difficult as it is often impossible to predict when the change in behavior will occur, and thus requires important decisions in terms of when to evaluate, how often to evaluate, and how to evaluate. Although it takes a greater effort to collect behavior change data, its value is important to training organizers as the data provides insight into the transfer of learning from the “classroom” to the work environment and whether or not behavior and practices have been changed in the direction intended by the training.

This Chapter discusses the importance of instructional design to facilitating behavior change and practice implementation post-training, outlines evaluation designs for the assessment of behavior change (including quasi-experimental and experimental approaches), identifies considerations in selecting an evaluation design, methods for measuring behavior and practice change (including observation and record or document reviews), and provides guidelines for accurately measuring behavior change and practice implementation.

SECTION 4.1

Evaluating Behavior and Practice Change Basics

The Domain of Measurement - What is Behavior Change?

Behavior evaluation in the training evaluation context examines the extent to which learning has been applied back on the job to change performance, practices and policies. Behavior evaluation is the extent to which trainees applied the learning and changed their behavior - this can occur immediately after and several months after the training. Behavior evaluation asks whether the relevant skills and knowledge trained on resulted in any measureable change in the activity and performance of the trainees when back in their roles. It can also ask if the change in behavior and new level of knowledge was sustained, whether the trainee was able to transfer his or her learning to another person, and whether the trainee is aware of his or her change in behavior, knowledge or skill level.

Instructional Design and the Evaluation of Behavior Change

The question posed at this level of evaluation checks on how the training affects job performance. Did the participants change their behavior based on what was learned? Has the job performance of the trainee improved because of the training? Just as a favorable reaction to a training program does not mean that any learning has occurred, learning the training program's material does not guarantee that the learning will be applied on the job and result in behavioral or practice changes. Kirkpatrick (1994) cites four conditions necessary for change to occur:

- The person must have a desire to change.
- The person must know what to do and how to do it.
- The person must work in the right climate.
- The person must be rewarded for changing.

Your instructional design can influence the first item (the desire to change) and provide the knowledge, skills, and attitudes needed for the second (knowing what to do and how to do it). However, the third and fourth items (the right climate and rewards for change) must be provided at the local level. Kirkpatrick (1994) cites five kinds of climates that may influence whether behavior change takes place:

- A preventing climate – In this climate, local management forbids the trainee to use or apply any of the training material to the job.
- A discouraging climate – In this climate, local management does not forbid application of the training to the job, but makes it obvious that changes in the way things are done are not desired.
- A neutral climate – In this climate, training is ignored.

CHAPTER FOUR: THE ASSESSMENT OF BEHAVIOR AND PRACTICE CHANGE

- An encouraging climate – In this climate, local management encourages individuals to learn and apply that learning to the job.
- A requiring climate – In this climate, local management knows what the individual has learned at a specific training and ensures that the learning is applied to the job.

As a training organizer you need to think seriously about what type of environment your trainees might be coming from and if you are providing them, through your instructional design, the skills or supports to overcome climates that are not conducive to the application of learned material and behavior change.

Ensuring that Your Instructional Design Facilitates Behavior and Practice Change:

- Provide opportunities to apply the skills learned through practical, hands-on activities
- Facilitate and foster organizational or management support for behavior change
- Ensure you are training on specific skills that are immediately applicable and relevant to trainee's jobs
- Implement follow-up procedures (action planning, "check-ins" to see how it is going)

Managing the Training Transfer – Facilitating Behavior Change

- Training should be linked to the specific performance objectives that relate to the needs of trainees. Revisit your training program logic model.
- Ideally, training participants' supervisors or individuals with decision-making authority and power to support and implement change should be involved in the development and delivery of training programs.
- Encourage pre-training briefings between training participants and their supervisors to share the training agenda and any pre-training materials. Pre-training briefings provide an opportunity to discuss what will be learned and how that will be applied when the trainee is back on the job.
- Encourage training participants to develop a learning action log at the end of the training and to share that log with their supervisors. This time can also be spent in further refining the action plan with the support and input of supervisors.
- Develop a peer or project team among the training participants to provide greater assistance with transferring the new skills and knowledge learned from the program back on the job.
- Set up a support portal on the web where additional and follow-up training materials and support for application of skills in the field are available.

Tip from Field-Test Sites

In order to build support for change in the workplace or field-setting, some of the sites involved in testing the recommendations of this Guide included key systems' leaders in training kick-offs to show their support. Leaders were also available throughout the training as well as after the training program in order to understand what participants had learned. Leaders were also selected for training advisory committee members to build buy-in for the behavior and practice changes desired by the training.

SECTION 4.2

Evaluation Designs for Measuring Behavior and Practice Change

Evaluation Designs for Behavior Change Assessment

To ensure that the behavior changed as a *result* of the training requires a degree of control in the research methodology. This can be achieved in one of two ways:

Quasi-Experimental Designs – One way to determine that behavior change is the result of the training is to conduct a quasi-experimental study. Quasi-experimental designs are those in which a treatment group or experimental group (i.e., the group that received the training) is compared to a control group (i.e., a group that did not receive the training). First, a similar group is selected for comparison. For example, the behavior of judges who did not go to the training can be compared to those who did attend the training. It will be important to select judges who are as similar as possible on relevant factors (i.e., the same jurisdiction with similar case loads and the same degree of prior experience and training). It will also be important to ensure the integrity of the comparison between the two groups by ensuring that the experimental group did not impact the control group in any way (i.e., that the judges receiving the training didn't share their resource materials and knowledge with the control group of judges).

Quasi-experimental designs may also use the same group for comparison purposes. In this situation, the same group is compared on pre-training behavioral assessments and post-training behavioral assessments. The same group of judges' behavior, for example is compared before and after the training with a pre- and post-test design. This will allow evaluators to determine specifically what changes have been made among a group that has been trained. (Please refer to Chapter Five for more detail about quasi-experimental designs.)

Experimental Designs – A very powerful method for examining behavior change is to use an experimental design. The experimental design is virtually identical to the quasi-experimental design, except for one key thing - experimental designs use random assignment to groups. This means, for example, that all judges in one jurisdiction would be randomly assigned to either attend the training or not to attend the training. In this situation, judges would be assigned to the training by virtue of chance alone. Once the random assignment is made, comparisons can be made between the two groups to see if their behavior is different. More detail about the experimental design approach to evaluation is provided in Section Six of this *Guide*.

Considerations in Selecting Behavior Change Evaluation Designs

Sampling – An important consideration when determining your behavior change assessment approach is the number of the participants you want to follow-up with to determine if behavior or practice changes have

CHAPTER FOUR: THE ASSESSMENT OF BEHAVIOR AND PRACTICE CHANGE

resulted after the training. Ideally, assessing 100% of participants would give you the best assessment. However, this is not always practical. Therefore, it may be more appropriate to use a *sample* (i.e., subsection) of participants. For best results, the sample should include as many participants as possible and as diverse a set of participants as possible to ensure that the assessment results would generalize to all participants. Revisit the discussion of sampling in Section Three of this *Guide*.

Consider cost versus benefits – The greater the potential benefits, the more time and money can be spent on the evaluation not only of behavior change but also of training impacts and results on organizational and systems change goals. Depending on what information you want to know, your available budget and the research method you employ, you can perform a variety of assessments.

- *Small scale.* Small scale behavioral assessments are ideal when there are budget constraints or appropriate research methodologies cannot be employed (e.g., quasi-experimental or experimental studies). An example of a small scale behavior change assessment would include administering surveys to training participants to self-report on any behavior or practice changes that have resulted since attending the training. For example, six months following training on “Allowing Children in the Courtroom,” a judicial officer might receive a questionnaire with the following questions.

How has your behavior on the bench changed as a result of the training?
How have you implemented what you have learned into your courtroom?
Have you found any barriers (i.e., resources) that exist which prohibit you from making change?

Of course, small scale self-assessments can be used with more advanced methodologies (i.e., quasi-experimental or experimental designs). However, by allowing participants to self-assess and not using more objective means of data collection, it is often difficult to ascertain the true level of behavior change.

- *Medium Scale.* Medium scale behavior change assessments are appropriate when more funding and research resources are available. In this case, self-report survey methods would be used but in a pre- and post-test design. Assessing behavior once before the training and once after the training would constitute a medium scale evaluation. In a medium scale assessment, pre-training behavior assessment would involve administering self-report questionnaires or behavioral assessment and practice ratings and checklists to training attendees before the training. In these instruments, trainees would be asked to describe current behaviors, activities and practices that are relevant to the training topic. Post-training, participants would be asked to describe their behaviors, activities and practices and any differences from pre-tests would be compared.

CHAPTER FOUR: THE ASSESSMENT OF BEHAVIOR AND PRACTICE CHANGE

- *Intermediate Scale.* At the intermediate level, self-report survey methods to assess behavior change would be augmented by other more objective measures of behavior change in a pre- and post-training quasi-experimental design. For example, behavioral observation and reports from supervisors, peers or colleagues about performance would be compared pre-training with behavioral observations and reports from supervisors, peers or colleagues about performance after the training.
- *High scale.* High scale evaluations are the most costly and require the most research assistance. High scale evaluations would constitute either advanced quasi-experimental design or experimental design. This might entail doing pre-training behavioral assessments with participants using multiple methods and conducting multiple follow-ups (for example, at 3, 6, and 12 months after the training). High scale evaluations may use a pre- and post-test comparison of the experimental (training) group as well as an assessment of a control group (the behavior of those who did not receive the training). In addition, random assignment to create the training and control group may be used to increase the confidence with which conclusions can be drawn about the training having caused behavior changes. These evaluations would provide the *best evidence* for behavior change resulting from training. Yet, they are also the most timely, difficult and costly to conduct.

When to Assess Behavior Change – Allow time for behavior change to take place. Although learning may be instant, at other times participants will want to consider the training more carefully. It may take some time for participants to be able to apply what they have learned and implement it into practice to make behavior changes. Further, resource constraints may make it difficult, or impossible, to adequately implement behavior change immediately. While you may have gauged participants' self-report or intent to change behavior immediately following the training, it's best to give participants at least two to three months following the training to begin to assess behavior change. Often, six months might be more realistic.

Summary Examples of Post-Training Behavior or Practice Change Evaluation Designs

Post-Training evaluation can be accomplished through some of these types of evaluation strategies:

I. Self-Reported Behavior Change

- Training participants report the intent to change behavior or practice immediately after training.
- Training participants report that behavior changed at X weeks after the end of training.

II. On- the-Job Behavior Change

- Trained individuals self-report that they changed their behavior or used the skill or knowledge on the job (within X weeks after the training).
- Trained individuals whose managers/ supervisors/ peers report that they changed their behavior/ used the skill or knowledge on the job after the training (with X weeks after the training).

CHAPTER FOUR: THE ASSESSMENT OF BEHAVIOR AND PRACTICE CHANGE

- Trained individuals that are actually observed to change their behavior/ use the skill or knowledge on the job after the training (within X weeks after the training).

III. On-the-Job Performance Change

- Trained individuals' managers/ supervisors/ peers report that their actual job performance changed as a result of their changed behavior / skill (within X weeks after the training).
- Trained individuals' managers/ supervisors/ peers report that their actual job performance changed as a result of their changed behavior / skill (within X weeks after the training).
- Trained individuals that have observable / measurable improvement (quality, timeliness, performance outcomes) in their actual job performance as a result of their changed behavior / skill (within X weeks after the training).
- Departmental or division performance compared to benchmarks (expressed in units with X% of individuals that went through the training). Assessment within X weeks after the training.
- Additional number of people who were trained (or cross-trained) by those who have previously attended the training and their change in skill/ behavior/ performance.

SECTION 4.3

Methods for Assessing Behavior and Practice Change

Several important techniques for measuring behavior change are described below (a number of these have already been covered in previous Chapters of the *Guide* as they are foundational tools for all levels of training evaluation measurement).

Survey and Interview Methods – Survey and interview individuals who can identify the behavior change you wish to see. This includes the participants, immediate supervisors or staff, peers or colleagues, and others who are knowledgeable about the behavior of interest. It is important to consider who is the best qualified to answer behavioral change or practice application questions, who is the most reliable, and who is the most available.

Example survey questions might include:

Please indicate your level of agreement to the following statements

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
In the last 6 months, the judge has discussed ICPC cases more at length	1	2	3	4	5
In the last 6 months, ICPC cases have moved more quickly through the court process	1	2	3	4	5

Observational Methods – Behavior on-the-job or in the field can be observed before the training and after the training and then compared. Observational techniques include providing supervisors, colleagues or peers with a structured assessment form that rates trainees on a number of skill sets, competencies, or specific behaviors relevant to the training topic.

One observational measurement technique is the 360 degree feedback approach (this approach was introduced in Section Four of the *Guide*). In the 360 degree feedback approach, trainees' behavior is evaluated from people who work around them (typically including managers, peers and subordinates). These individuals fill out an anonymous feedback form (often online) that asks questions covering the behaviors or practices of interest to the training organizers. The feedback forms include questions that are measured on a rating scale and also ask raters to provide written comments. The person receiving feedback also fills out a self-rating survey that includes the same survey questions that others receive in their forms. The 360 degree feedback approach not only provides valuable evaluation data for training organizers but also serves as an additional learning opportunity for training participants as it gives them a clear picture of their greatest overall strengths and weaknesses. Individual responses should be combined with responses from other people in the same

CHAPTER FOUR: THE ASSESSMENT OF BEHAVIOR AND PRACTICE CHANGE

rater category (e.g. peer responses should be combined, subordinates' responses combined, etc.) in order to preserve anonymity and to give the trainees a clear picture of his/her greatest overall strengths and weaknesses.

An additional observation technique (which may be less threatening or uncomfortable for training participants) is structured field-observation. In structured field-observation, evaluators use standardized observation tools to code behaviors in the field. For example, a court observation instrument can be used to code child abuse and neglect hearing practice prior to the training. The same instrument can then be used after the training to collect information about hearing practices, with particular attention to procedures and behaviors trained on. Any pre-training and post-training observations can be compared to determine if the training resulted in any changes. If training was conducted with judges that included a focus on making clear, verbal reasonable efforts findings verbally in hearings, for example, a court observation instrument might include the following coding questions to determine if the judges are applying what they learned:

Hearing Type: _____				
Hearing Date: ____/____/____				
Were verbal reasonable efforts findings made? <input type="checkbox"/> Yes <input type="checkbox"/> No				
Indicate the specificity of the reasonable efforts findings:				
Not at all Specific	Somewhat Specific	Very Specific		
1	2	3	4	5

Field-observations should be conducted by neutral observers who fill out forms designed to capture the behaviors of interest. A neutral observer is an individual experienced in the dependency court system and process who is not directly involved in the training. In the case of court observation, neutral observers are observers who are not involved in the observed case or courtroom. The aim is to have a knowledgeable observer who is unbiased in his or her coding of the event under observation.

- **The Accuracy of Observation and Check-Coding** – The issue of observer bias or inaccuracy can be addressed in a number of ways. One approach is through rigorous training of several observers on the observation instrument and in the actual field-setting or event that will be coded for the appearance of specific behaviors or practices. After training, a sample of events should be coded by more than one observer in order to pre-test the observation instrument and the reliability of the coding strategy. Discrepancies between the results of coders who code the same event can indicate coder-error or problems with the instrument and coding process itself. Pre-testing of the observation instrument is essential to make any modifications necessary to ensure the accuracy of the evaluation process. In addition, it is ideal to use more than one coder of the same event if possible, at least until you are assured of the accuracy of the observation instrument in reliably capturing an observed event. When using more than one coder, individual coders do not refer to each other as they are coding an event.

CHAPTER FOUR: THE ASSESSMENT OF BEHAVIOR AND PRACTICE CHANGE

After coding, each individual's observations of behavior are compared to the observations of others to establish inter-rater reliability. Your goal is to obtain high inter-rater reliability (or high agreement between coders).

“Check-coding” procedures may also be implemented. To ensure accuracy, a small sample of events is “check-coded” (double-checked) by an experienced coder who notes any discrepancies in the way the event was coded. All errors are then addressed in consultation with the original coder. If a high rate of error occurs in the sample that is check-coded, those observations should be discarded for analysis purposes.

- **The Observation Effect** – Behaviors may change in the direction of what individuals were trained to do merely as a result of being observed - “the observation effect.” This is also known as the “Hawthorne effect”⁶⁵ - a form of reactivity whereby individuals improve or modify an aspect of their behavior that is being evaluated or measured simply in response to the fact that they are being studied. However, research has demonstrated that if observation is unobtrusive, and conducted over sufficient time, people being observed forget that they are being observed and behave naturally.⁶⁶ When conducting field-observations of behavior, therefore, it is important to implement frequent and regular observations.

Secondary Sources – Secondary data can be defined as “information gathered for the purpose other than the immediate or first application.”⁶⁷ Relevant secondary data sources for the dependency court context include existing evaluation or performance studies and administrative records and reports (e.g., management information system data; case files; performance data). Training evaluation studies have infrequently used these data sources, but their utility to the assessment of behavior change and training impact evaluation is undeniable. One example of a secondary data source that is especially valuable in dependency court evaluation contexts is the case file.

Depending upon the behavior or practice change of interest, an examination of case files or other records or relevant documents may be an extremely valuable measurement. For example, if the goal of training was to impact specific behaviors in court hearings, case files before the training can be examined to determine if those behaviors are clearly present and able to be coded from documents that summarize the activities of the hearings. For example, for a training on strategies to improve educational outcomes for youth in foster care, case files may be examined pre-training to determine the extent of active inquiry from the bench in hearings about the educational achievement and progress of foster youth and whether educational services are ordered. Post-training, follow-up case file review can ascertain whether changes in these behaviors have resulted as compared to pre-training file review (e.g., court orders can be examined for details about services).

⁶⁵ Adair et al., (1988).

⁶⁶ Shayer, M. (1992).

⁶⁷ Rossi, Freeman & Lipsey (1999).

CHAPTER FOUR: THE ASSESSMENT OF BEHAVIOR AND PRACTICE CHANGE

The training and check-coding procedures outlined above in relation to structured field-observation techniques also apply to case file review (e.g., a sample of coded files should be coded by an additional, experienced coder to determine coding error).

As an example, judicial officers might attend training on Making Reasonable Efforts. To determine what impact they had, a case file review form might include the following items:

Hearing Type: _____ Date: _____			
Were reasonable efforts findings made orally in court? <input type="checkbox"/> Yes <input type="checkbox"/> No			
Indicate the degree of discussion regarding reasonable efforts in the hearing:			
Not at all	Minimal	Somewhat	Substantial
1	2	3	4

Evaluating secondary data can be more complex than at the reaction or learning level. Nevertheless, it is important and should be built into any training evaluation. Surveys and interviews work well at this level, as well as direct observation, something fairly easily accomplished. In some cases, review of secondary data sources such as case files may be necessary to increase the objectivity of the measurement. Ideally, statistical comparisons of an individual's behavior or on-the-job performance both before and after training should be conducted.

Summarizing an Approach to Measure Behavior or Practice Change

Basic Questions – Do training participants remember and understand how to use the skills, knowledge and attitudes that they were taught? Do they successfully acquire skills, knowledge or attitudes trained on and applied on the job? Do training participants change their behavior and practice as a result of the training?

When to Measure – Evaluation should take place a sufficient amount of time after the end of the training to ensure that participants have had the opportunity to use and practice learned skills in the workplace and change their behavior. Consider measurements at three months up to two years from the training completion. Measurements should not be done more than two years after the training as participants may have difficulty remembering how the training affected their work. In addition, practice standards and other environmental or organizational changes may have taken place that affect the ability to apply the behavior or practice (irrespective of the training

How to Measure:

- Interviews or surveys of training participants, their supervisors, peers and other that can report on workplace or field-setting behaviors.
- Observation of training participants in their workplace or field-setting.

CHAPTER FOUR: THE ASSESSMENT OF BEHAVIOR AND PRACTICE CHANGE

- Performance data (i.e., if participants were trained to increase the frequency with which they ask about the educational needs of children in foster care during dependency court hearings, then the frequency with which they inquire about educational needs post-training can be measured and compared to pre-training performance on this issue).
- Secondary data sources (i.e., examine court case files both pre- and post-training to determine the frequency with which the practices trained on appear in documents).

Summarizing an Approach to Measure Behavior or Practice Change

Basic Questions – Do training participants remember and understand how to use the skills, knowledge and attitudes that they were taught? Do they successfully acquire skills, knowledge or attitudes trained on and applied on the job? Do training participants change their behavior and practice as a result of the training?

SECTION 4.4

Guidelines for the Measurement of Behavior and Practice Change

The following guidelines are offered to assist in the implementation of behavior and practice change evaluation measurement.

- Use a control group whenever possible
- Give the trainee time to review the training material and apply it to his/her job. Unless the training is provided to implement a dramatic change in the way things are done, the trainee will likely experiment with any changes prior to full implementation of those changes.
- Evaluate behavior and job performance both before and after the training. Through this comparison any change can be observed and the change attributed to training.
- Consider carefully who should be interviewed or surveyed in order to determine behavior changes. Who is in the best position to see any change that occurs? It may be the trainee, his/her immediate supervisor or subordinates; or it may be the trainee's peers. As a decision is made about who to select for the evaluation, four questions should be considered:
 - Who is best qualified to observe the change?
 - Who is most reliable in providing honest responses?
 - Who is most available to answer questions or complete a survey?
 - Are there any reasons why one of the four people mentioned above should not be used?
- Determine whether you need to measure behavior changes in all trainees or just focus on a sample. The answer to this question may depend upon the size of the group who are trained. In general, the more people evaluated the more confidence you can have in any conclusions drawn.
- It may be necessary to repeat a behavior level evaluation several times in order to see the full impact of the change. For example, a series of evaluation surveys or observations at three, six, and nine months after the training might be appropriate in some situations. It is good to remember, however, that factors, other than the training, can have an impact on job performance if the evaluation is conducted too long after the training itself.
- It may be important to consider the cost of the evaluation.

TOOLS AND RESOURCES – CHAPTER FOUR

- Behavioral Observation Tool for Supervisors or Mentors
- Court Hearing Observation Form
- Sample Behavioral Self-Assessment
- Sample Behavioral Self-Assessment Survey Form
- Sample Case File Review Form

CHAPTER FOUR: References and Resources

- Adair, J.G., Sharpe, D. & Huynh, C.L. (1989). Hawthorne control procedures in educational experiments: A reconsideration of their use and effectiveness, *University of Manitoba Education Review*, Vol 59 (2), 215-213.
- Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Review*, 26(3), 201-213.
- Bramley, P. (1996). *Evaluating training effectiveness*. Maidenhead: McGraw-Hill.
- Bernthal, P. R. (1995). Evaluation that goes the distance. *Training and Development Journal*, 49(9), 41-45.
- Bushnell, D. S. (March, 1990). Input, process, output: A model for evaluating training. *Training and Development Journal*, 44(3), 41-43.
- Dixon, N. M. (1996). New routes to evaluation. *Training and Development*, 50(5), 82-86.
- Donoghue, F. (1999). *Promoting added value through evaluation of training*. Dublin: European Commission Leonardo-PAVE Project.
- Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. *Journal of the American Society of Training Directors*, 13, 3-26.
- Kirkpatrick, D.L. (1994). *Evaluating training programs, the four levels*. Berret-Koehler Publishers, San Francisco,.
- McEvoy, G. M., & Buller, P. F. (August, 1990). Five uneasy pieces in the training evaluation puzzle. *Training and Development Journal*, 44(8), 39-42.
- Phillips, J. J. (1991). *Handbook of training evaluation and measurement methods*. (2nd ed.). Houston, TX: Gulf.
- Rossi, P.H., Freeman, H.E. & Lipsey, M.W. (1999). *Evaluation: A systematic approach* (6th ed.). Thousand Oaks: Sage.
- Rossi, P.H., Freeman, H. E., & Wright, S. R. (1979). *Evaluation: A systematic approach*. Beverly Hills, CA: Sage.
- Shayer, M. (1992). "Problems and issues in intervention studies." In Demetriou, A.; Shayer, M.; Efklides, A. *Neo-piagetian theories of cognitive development: Implications and applications for education*. London: Routledge, pp. 107–121.

CHAPTER FIVE

The Assessment of Training Outcomes

The level of evaluation that measures the success of the training program in achieving organizational and system goals, such as improved performance, improved quality, improved collaboration, and, in the dependency court system context, improved safety, and timeliness, due process, permanency, and well-being outcomes is frequently thought of as the ultimate bottom line. From a court improvement and systems change perspective, this is ultimately the reason for designing and implementing dependency court training programs. Yet this level of evaluation - the outcomes level - is typically not addressed in training evaluations. This is due to the fact that determining the outcomes of trainings is more difficult to measure, and it is hard to link those results directly with the training itself (this is especially true for trainings in complex systems such as the dependency court system).

This Chapter defines outcome measurement and outcome evaluation terminology, summarizes common research models for outcome evaluation, provides a process for determining the best evaluation approach, and reviews data collection methods for outcome measurement. The Chapter concludes with recommendations for excellence in training evaluation to facilitate an outcome-focus.

SECTION 5.1

Evaluating Training Outcome Basics

The Outcome Domain of Measurement

The final domain of measurement is the evaluation of the outcomes that occur as a result of your training program. This level of measurement determines the training program's effectiveness, that is, it answers the question, "What outcomes has the training achieved?" Outcome areas of interest in the dependency court context may include innovation, collaboration, safety, timeliness, due process, permanency and well-being. Collecting, organizing and analyzing information at the outcome level of evaluation can be difficult, time-consuming and more costly than the other three levels, but the results are often worthwhile when viewed in the full context of their value to the organization, and ultimately to the dependency court system. As mentioned in earlier sections of this Guide, perhaps the most frequent type of training evaluation measurement is focused on Level One, Satisfaction Measurement, because that is the easiest level to measure. However, in many ways, it provides the least valuable data. Measuring results that affect the organization is considerably more difficult, but this level of measurement yields the most valuable information.

Ideally, each evaluation level should be used to provide multiple datasets for measuring the training program's effectiveness.

- Satisfaction or reaction data inform you about how relevant the training is to the work the learners perform (it measures how well the training requirement analysis processes worked).
- Learning informs you about whether the training content effectively transferred relevant knowledge, skills and attitudes to the learners (it measures how well the training design, development processes and implementation processes worked to transfer learning).
- Behavior Change or performance data inform you about the degree to which the learning is actually applied to the learner's job and changes in practice are made.
- Outcome data inform you of the "return" that individuals, the organization, and system receive from the training.

The first three levels (satisfaction, learning, and behavior change) give you "information" for improving the learning package or training program, while the fourth-level (outcomes) gives you "impacts." Instead of evaluating how well the training "informed", it evaluates the impact the training has had on individuals, organizations, and systems.

CHAPTER FIVE: THE ASSESSMENT OF TRAINING OUTCOMES

Basic Terminology

Below we review some basic terminology that has relevance to outcome evaluation; we also introduce some new terms.

- **Inputs** – These are the materials and resources that the training program uses in its activities, or processes (e.g., equipment, staff, volunteers, facilities, money, etc.). These are often easy to identify and many of the inputs seem common to many organizations and training programs.
- **Activities** – These are the projects, or processes, that the training program undertakes (e.g., designs a curriculum, implements workshops, etc.).
- **Outputs** – These are the units of service, for example, the number of people taught, the number of advisory group meetings held, the number of participants trained, the number of materials produced. This information, while providing valuable data about program implementation indicates nothing about the actual impacts/benefits/changes in participants who went through the training program – the number of participants trained merely indicates the headcount of those who went through your training program.
- **Outcomes** – These are the actual impacts/benefits/changes that result from your training program. These outcomes are usually expressed in terms of: knowledge and skills (these are often considered to be rather short-term outcomes); behaviors (these are often considered to be rather intermediate-term outcomes); and values and conditions (these are often considered to be rather long-term outcomes). An outcome evaluation of your training program investigates whether the training resulted in demonstrable effects on specifically defined target outcomes.

Outcomes are the observable behavioral, organizational and system changes that take place as a result of the training. They are the observable positive or negative changes in the actions of individuals that have been influenced, directly or indirectly, partially or totally, intentionally or not, by your training activities or your “outputs” that potentially contribute to the improvements envisioned in by you in your vision for the training.

- **Outcome targets** – These are the number and percent of participants that you want to achieve the outcome, for example, an outcome goal of 95% of social service case plans will be filed within the statutorily mandated timeframe for the filing of social service case plans.

- **Outputs** (hardly indicate anything about changes - they're usually just numbers)
- **Outcomes** (indicate true changes)
- **Outcome targets** (specify how much of your outcome you hope to achieve)
- **Outcome indicators** (what you can see, hear, read, etc., and that suggest you are making progress toward your outcome target or not)

CHAPTER FIVE: THE ASSESSMENT OF TRAINING OUTCOMES

- **Outcome indicators** – These are observable and measurable “milestones” toward an outcome target. These are what you'd see, hear, read, etc., that would demonstrate to you whether you're making any progress toward your outcome target or not. For example, the percent of case plans that are compliant with statutorily mandated time frames for filing six months after a training program designed to address the timeliness of case plans. This indicator would give you a strong impression as to whether or not 95% of social service plans filed will be filed within the statutorily mandated timeframe over the next year.

Outcome: Fathers will improve their parenting knowledge and skills

Outcome target: XXXXXXXXXXXX

Outcome Indicator: Statistically significant increase between pre- and post-test average scores on parenting inventory

Summative and Formative Evaluation

When discussing outcome evaluation, it is important to distinguish it from what it is not - There are many different types of evaluations depending on the object being evaluated and the purpose of the evaluation. Perhaps the most important basic distinction in evaluation types is that between *formative* and *summative* evaluation. Formative evaluations strengthen or improve the object being evaluated. They help form it by examining the delivery of the program, the quality of its implementation, and the assessment of the organizational context, personnel, procedures, inputs, and so on. Summative evaluations, in contrast, examine the effects or outcomes of some object. They summarize it by describing what happens subsequent to delivery of the program; assessing whether the program can be said to have caused the outcome; determining the overall impact of the causal factor beyond only the immediate target outcomes; and, estimating the relative costs associated with the training.

Formative evaluation includes several evaluation types:⁶⁸

- **Needs assessment** determines who needs the program, how great the need is, and what might work to meet the need.
- **Evaluability assessment** determines whether an evaluation is feasible and how stakeholders can help shape its usefulness.
- **Structured conceptualization** helps stakeholders define the program, the target population and the possible outcomes.
- **Implementation evaluation** monitors the fidelity of the program.

⁶⁸ Clearly, this listing is not meant to be exhaustive. Each of these methods, and the many not mentioned, is supported by an extensive methodological research literature.

CHAPTER FIVE: THE ASSESSMENT OF TRAINING OUTCOMES

- **Process evaluation** investigates the process of delivering the program, including alternative delivery procedures. Process evaluation is aimed at describing the course and context of a program or intervention. This type of evaluation helps in the interpretation of impact/ outcome evaluation findings and can be fed into formative evaluation to improve the program. It can also be used to identify best practices, which can be used to improve other programs in the future.

In **formative research** the major questions and methodologies are:

- **What is the definition and scope of the problem or issue, or what's the question?**
 - Formulating and conceptualizing methods might be used including brainstorming, focus groups or Delphi methods⁶⁹, and stakeholder surveys.
- **Where is the problem and how big or serious is it?**
 - The most common method used here is "needs assessment" which can include analysis of existing data sources, and the use of sample surveys, interviews of constituent populations, qualitative research, and focus groups.
- **How should the program be delivered to address the problem?**
 - Some of the methods already listed apply here, as do project planning and implementation methods like flow charting.
- **How well is the program or technology delivered?**
 - Qualitative and quantitative monitoring techniques, the use of management information systems, and implementation assessment would be appropriate methodologies here.

Summative evaluation can also be subdivided:

- **Outcome evaluations** investigate whether the program caused demonstrable effects on specifically defined target outcomes.
- **Impact evaluation**, while related to outcome evaluation, is actually broader and assesses the overall or net effects – intended or unintended – of the program as a whole. Impacts can be thought of as the positive and negative, primary and secondary **long-term effects** produced by your training intervention - directly or indirectly, intended or unintended. Impacts are the significant, structural, sustained and positive improvements resulting from the policy and practice changes that resulted from your training program.
- **Cost-effectiveness and cost-benefit analysis** address questions of efficiency by standardizing outcomes in terms of their dollar costs and values.
- **Secondary analysis** re-examines existing data to address new questions or use methods not previously employed.

⁶⁹ The Delphi method is characterized as a structured method for group communication that is effective in allowing a group of individuals, as a whole, to deal with complex problems (Linstone & Turoff, 2002).

CHAPTER FIVE: THE ASSESSMENT OF TRAINING OUTCOMES

- **Meta-analysis** integrates the outcome estimates from multiple studies to arrive at an overall or summary judgment on an evaluation question.

In **summative research** the major questions and methodologies are:

- **What type of evaluation is feasible?**
 - Evaluability assessment can be used here, as well as standard approaches for selecting an appropriate evaluation design (see later chapters in this section of the Guide).
- **What was the effectiveness of the program?**
 - One would choose from observational and correlational methods for demonstrating whether desired effects occurred, and from quasi-experimental and experimental designs for determining whether observed effects can reasonably be attributed to the intervention and not to other sources.
- **What is the net impact of the program?**
 - Econometric methods⁷⁰ for assessing cost effectiveness and cost/benefits would apply here, along with qualitative methods that enable us to summarize the full range of intended and unintended impacts.

Outcome evaluation of training programs is facilitated by:*

- Leadership support for the evaluation - visible support for the outcome evaluation approach helps to provide access to data sources, encourages participation in data collection procedures, and builds support for using the findings
- Commitment of time and staff resources - outcome evaluation can be especially time and staff intensive as it introduces additional measurement procedures and extends the overall time needed for data collection
- Program stability - training programs that are too new with untested faculty and curricula are not good candidates for introducing outcome measurement
- Evaluation expertise - individuals with knowledge of evaluation methods and analytic procedures may be necessary when implementing research designs necessary for outcome evaluation
- Management information systems and other data sources with the capability to generate key performance data (a complete sentence is needed to make the format of this item similar to the format of the other items above)

⁷⁰ "The econometric approach develops explicit models of outcomes where the causes of effects are investigated and the mechanisms governing the choice of treatment are analyzed," (Heckman, 2008).

SECTION 5.2

Strategies for Facilitating Outcome Measurement Evaluation

Establish an Outcome Measurement Working Group

Because outcome measurement can be the most difficult of the levels of training evaluation, you should plan your training with an eye to facilitating outcome measurement. For example, one strategy you may want to implement in order to facilitate outcome measurement is to establish an outcome measurement working group. This working group, consisting of training managers and evaluation consultants, can help work out the details of the outcome measurement process and oversee its implementation. Once this group has developed a framework for the evaluation, additional stakeholders should be invited to provide feedback on the measurement plan. Including representatives from different parts of the dependency court system in this input or feedback process can provide a rich perspective on what outcomes should be measured, how outcome information can be collected, and the ways outcome information can be used. Including input from stakeholders to help develop and implement your outcome measurement process will also help to build “buy-in” for the assessment and generate commitments for accessing needed data sources. Be sure to establish a timeline or schedule for the working group that allows enough time to work through all of the issues that may arise. You will probably need a number of meetings to reach consensus on a measurement plan and data collection procedures.

Examine your Training Program’s Mission, Goals, and Objectives to define your Outcomes

In planning for an outcome measurement process, it is important to re-visit your training program’s mission, goals, and objectives. Next, the objectives statements should be translated into intended or desired training outcomes or results. These should be as specific as possible as they will become the basis for identifying specific outcome indicators.

As other sections of this Guide have discussed, to help your working group identify specific training outcomes some of the items you can look at are:

- Outcomes identified by similar training programs.
- Outcomes identified by previous evaluation reports.
- Meaningful and relevant performance measures such as safety, permanency, timeliness, and due process and well-being measures (e.g., national performance dependency court system performance measurement standards).
- Focus group data from stakeholders that have identified their expectations for trainings or what they have considered a successful program.

CHAPTER FIVE: THE ASSESSMENT OF TRAINING OUTCOMES

- Use your training program logic model which outlines the steps from inputs to activities to outputs, which ultimately lead to the outcomes that are expected to result.

Application of the outcome level of measurement to your training evaluation will require you to either develop or re-visit your training program logic model (Refer to Chapter 1.7 for an overview of logic models). What were the desired impacts or outcomes defined at the beginning of the training design? For example, your logic model may have identified greater organizational efficiency, a change in the way a target stakeholder group behaves, or improvements in specific performance outcomes such as permanency outcomes as desired impacts of the training program. In addition, you should ask yourself if there have been other elements, besides your training program, that may have influenced the outcome you want to measure. Your training program or training “intervention” cannot be viewed in isolation; there may have been other factors influencing the changes you observed (e.g., the infusion of additional staff resources, a change in the law, etc.). Identifying these factors will help you to assess the level of influence your training program has had on outcomes compared to other factors.

Below is a list of some of the impacts or outcomes you may have considered in your logic modeling process:

- **Organizational Outcomes** – Group operation and management improvements; technical operation and management improvements; institutionalization of change practices; diffusion of training concepts to new groups of individuals; new linkages with system partners; stronger and more diverse collaboration; improvement in stakeholder performance; new service deliveries, methods and strategies designed and implemented; new philosophy, purposes, and goals; improved organizational culture.
- **Dependency Court System Performance Improvement** – Improved Safety, Timeliness, Due Process, Permanency and Well-Being for Children and Families.
- **Individual Outcomes** – Changes in knowledge, attitudes, skills; sustainable practice; change in aspirations, self-image, perspectives; use of methods, services, strategies; invention of new methods, services, strategies; compliance with standards for practice; improved patterns of communication and working relationships.
- **Community Outcomes** – Changes in the administration of justice and child welfare; changes in public opinion about the effectiveness of the dependency court system; changes in court-agency-community communication; improved access to and coordination of community resources for children and families involved in the dependency court system; evidence of cultural practice changes.

The training logic model you designed during your training program development should have outlined appropriate outcomes that your training program hopes to achieve.

CHAPTER FIVE: THE ASSESSMENT OF TRAINING OUTCOMES

Table 5-1
Example of a basic logic model

Contextual Factors	Training Activities	Individual Outcomes	Organizational/System Outcomes	Long-Term Outcomes
Nature of collaboration court-agency-community; Level of personnel involved; Organizational constraints; Training policies; CFSRs	Develop curriculum Deliver Training	Stakeholder Skills <ul style="list-style-type: none">Improved KnowledgeImplementation of New PracticeImproved Behavior/Performance	Improved Child Abuse and Neglect Case Practice/Performance <ul style="list-style-type: none">TimelinessDue ProcessSafetyPermanency	Improved Child and Family Well-Being

True outcome or impact evaluation requires a quasi-experimental or experimental design [more about these designs in Chapter Five of this Section], and typically involves:

- Use of a control or comparison groups (preferably with random assignment to groups but if not, ensure that comparison groups are as similar as possible to your intervention or training group)
- Use of pre- and post-test measures (check consistency in use of hyphen throughout document)
- Use of multiple data sources and multiple methods (both qualitative and quantitative)
- Use of fidelity checks (do control and comparison groups remain intact? Are interventions that the training group received implemented in the field as intended?)

Determine the Best Outcome Measurement Approach

Evaluating impacts and outcomes is the most crucial of the four levels of evaluation measurement. When evaluating impacts, you want to use similar guidelines as those first introduced in Chapter Three (learning acquisition measurement) and then further discussed in Chapter Four on measuring behavior change. First, use a control group or comparison group (even if your comparison group is the same individuals measured before and after the training). [More detail about research designs for outcome measurement in Section 5.3]. Allow appropriate time for the new knowledge, skills and attitudes and changes that you were attempting to address with the training to take effect. Learning and behavioral change takes different amounts of time for different individuals. The level of learning and behavioral change can also depend on the subject and certain situations. Some things take longer to see results. Changes may occur right away because of input from management, or excitement to try something new. Consider repeating your measurement at different times to ensure retention and determine whether changes are sustained.

CHAPTER FIVE: THE ASSESSMENT OF TRAINING OUTCOMES

Select Specific Indicators and Data Sources to Measure your Outcomes

After you have defined your training program outcomes, your working group should translate those outcomes into specific indicators that will be measured. For each outcome, identify one or more outcome indicators. Outcome indicators should almost always use words such as “the number of ...” or “the percent of ...” Before you finalize your indicators, be sure to think about how you would go about collecting data for that indicator.

Data sources and collection procedures need to be carefully thought out in order to provide reliable and valid data on your outcome indicators. Advice from someone who is experienced in outcome measurement or evaluation methods is helpful when you begin to identify data sources for your outcome indicators. If you do not have an internal capacity, you might consider finding an expert from a local university or community college, or seek a volunteer expert or consultant. See Section 5.4 for more information about possible valuable data sources for outcome measurement.

Identify Key Training Participants’ Characteristics to Link to Outcomes

Your outcome evaluation working group should identify key training participant characteristics to link to outcome measures. This information will enable you to assess the extent to which the training program has been, successful in helping specific groups of participants acquire new skills, change behaviors or implement a new practice. Some of the key training participant characteristics to consider are gender, race/ethnicity, professional role, years of experience, county or jurisdiction, etc.

Pre-Test your Outcome Evaluation Procedures, Make Needed Modifications, and Implement

Any new data collection procedures such as surveys or observation protocols should be pre-tested. For example, provide a copy of your survey to a sample of individuals who represent the audience for your training. Have these individuals complete the survey and provide you with feedback about whether they understood the questions. Use their feedback to make any necessary modifications to the survey questions, ordering, etc. In addition, run analyses based on the completed surveys in your pre-test sample. Are the findings generated useful to you? Use this information to eliminate questions that are not providing useful or valuable information.

CHAPTER FIVE: THE ASSESSMENT OF TRAINING OUTCOMES

To pre-test document or record review and observation methods, have at least two different individuals use the same coding protocol to code the same document or observe the same event. Examine and discuss any differences between coders. Ask yourself if the differences are due to inaccuracies in the coder's application of the code. If they are, then you will need to re-train on the coding procedures. However, if any differences are

Ethics: Informed Consent and IRB Approval

If you plan to include in your evaluation the focus and reporting on personal information, then you should first gain individuals' consent to do so. Your evaluation subjects should understand what you're doing with them in the evaluation and how any information associated with them will be reported. You should clearly convey terms of confidentiality regarding access to evaluation results. They should have the right to participate or not. Have participants reviewed and sign an informed consent form (see the Section Five Tools and Resources for an example). In addition, some data collection methods, such as those requiring access to child welfare, school or health records or involving interviews with children, youth or parents involved in the dependency court system, may require the prior approval of a "Human Subjects" or Institutional Review Board (IRB). For assistance with knowing when an IRB approval is needed and the materials required for submission to an IRB, see www.hhs.gov/ohrp/index.html

due to the code itself, you will need to re-visit your coding protocol and make refinements (e.g., the coding protocol is too ambiguous and requires too much subjective interpretation or judgment-calls, which introduces error). As with surveys, be sure to run analyses on completed record review and observation instruments to determine if the findings generated are meaningful. Use that information to eliminate any superfluous coding questions.

Develop an Analysis Plan

After outcome indicators and data collection methods have been selected, it is helpful to outline a plan for analysis. Your data analysis plan clearly identifies the specific techniques you will use to examine the data generated by your evaluation research method and instrumentation. It is a road map for generating and reporting findings. While the actual analysis will not be conducted until data are generated, preparing an analysis plan helps to make the analysis process flow more smoothly and helps to produce a timely evaluation research report. In addition, sharing the analysis plan can obtain buy-in from key stakeholders for access to data sources and, ultimately, for the final conclusions drawn about the success of the training program.

Typically, you develop an analysis plan in parallel with your data collection instruments. The analysis plan is tied back to the goals and objectives of your training evaluation. In addition to the obvious purpose of an analysis plan, producing a plan serves to improve the instrumentation development process and to manage project scope.

CHAPTER FIVE: THE ASSESSMENT OF TRAINING OUTCOMES

In preparing your analysis plan, consider these steps:

- State the key evaluation objectives clearly at the beginning of the analysis plan and refer to them throughout the process.
- Describe the major comparisons for the analysis (e.g., major cross tabulations for the evaluation such as: judges with 0-3 years of previous dependency court experience versus judges with more than 3 years of dependency court experience; participants that are “satisfied” with the training overall versus participants that are “dissatisfied”).
- Think through how you expect to present the results from each of your evaluation research questions. What statistics, if any, will you use in the analysis?
- When the analysis plan is finished, go back and make sure each key evaluation objective has been addressed.

The key is to focus on your objectives and think critically about how to execute on the primary goals of your evaluation study.

See the Chapter Five Tools and Resources for a list of the key steps in planning an outcome-focused training evaluation, examples of outcomes and indicators, and tips for identifying outcomes, data sources and data collection methods.

SECTION 5.3

Major Models of Outcome Assessment

Quasi-experimental and experimental evaluation designs are ones that permit outcome and impact assessment. The purpose of these approaches is to determine whether changes in outcomes were due to the contributions of the training program and not just to life's experiences, other influences or some other interventions (e.g., such as a change in the law, the influx of additional resources, media attention to a specific issue, transitions in personnel, the implementation of a new policy unrelated to the training program, etc.). Quasi-experimental and experimental models ask the question, "Were differences in outcomes of interest attributable to the training program?" The simplest way to determine causality between the training program's inputs, activities, and comparable groups is to examine outcomes for a group that received the educational treatment and a group that did not. This means that training program accessibility, at least during the experimental phase, must be withheld from the control or comparison group. In the experimental model, assignment to the "treatment" or training group and control group would also be random, increasing the confidence with which any differences you find between the groups can be attributed to the training intervention. In addition, when using comparison groups, you must ensure fidelity to the research design. "Fidelity checks" involve both determining whether the groups remained intact (did the training group inadvertently make the knowledge, skills or strategies learned in the training available to the comparison group?) and whether the knowledge, skills or strategies learned were implemented as intended in the field by the experimental group (fidelity to the 'model'). More detail about quasi-experimental and experimental designs for training evaluation outcome measurement is offered below.

True experimental design

In a simple case of a true experimental design, members of a group (e.g., dependency system stakeholders) are identified for study. These are the individuals on which the intervention (your training program) is going to focus. A sample of individuals is taken from this group and randomly divided into two groups - one half of the group of individuals will have the training program applied to them (the **intervention, experimental or treatment group**) and the other half of the group of individuals will not have the training program applied to them (the **control group**). Changes in measurements of the outcomes of interest you have identified are then compared before and after the training intervention. It is presumed that any difference (beyond what is estimated as likely to have occurred as a result of chance), has been caused by the intervention. This is because there is no reason to believe that the individuals in the intervention and the control group differ in any systematic ways. Therefore it is presumed that what created the difference in results between the intervention and control groups is the fact that individuals either received or did not receive the intervention or training program.

A variation on the basic true experimental design is called a '**waiting list**' or '**pipeline**' design. This design uses

CHAPTER FIVE: THE ASSESSMENT OF TRAINING OUTCOMES

the same approach as the true experimental design; however the intervention or training program is only withheld from those in the control group for a limited period of time (the time they spend on a 'waiting list' of some sort). This is in contrast to a true experiment where the control group would never receive the intervention. This design is often regarded as more appropriate (because the control group does not miss out on the intervention) and more feasible (because participants and stakeholders are more likely to accept it and agree to participate) than with true experiments. A problem with the design however, is that effect of the intervention or training program truly has to be withheld from the 'waiting list' group (i.e., you have to ensure that stakeholders who have already participated in the training do not share their knowledge with those on the waiting list).

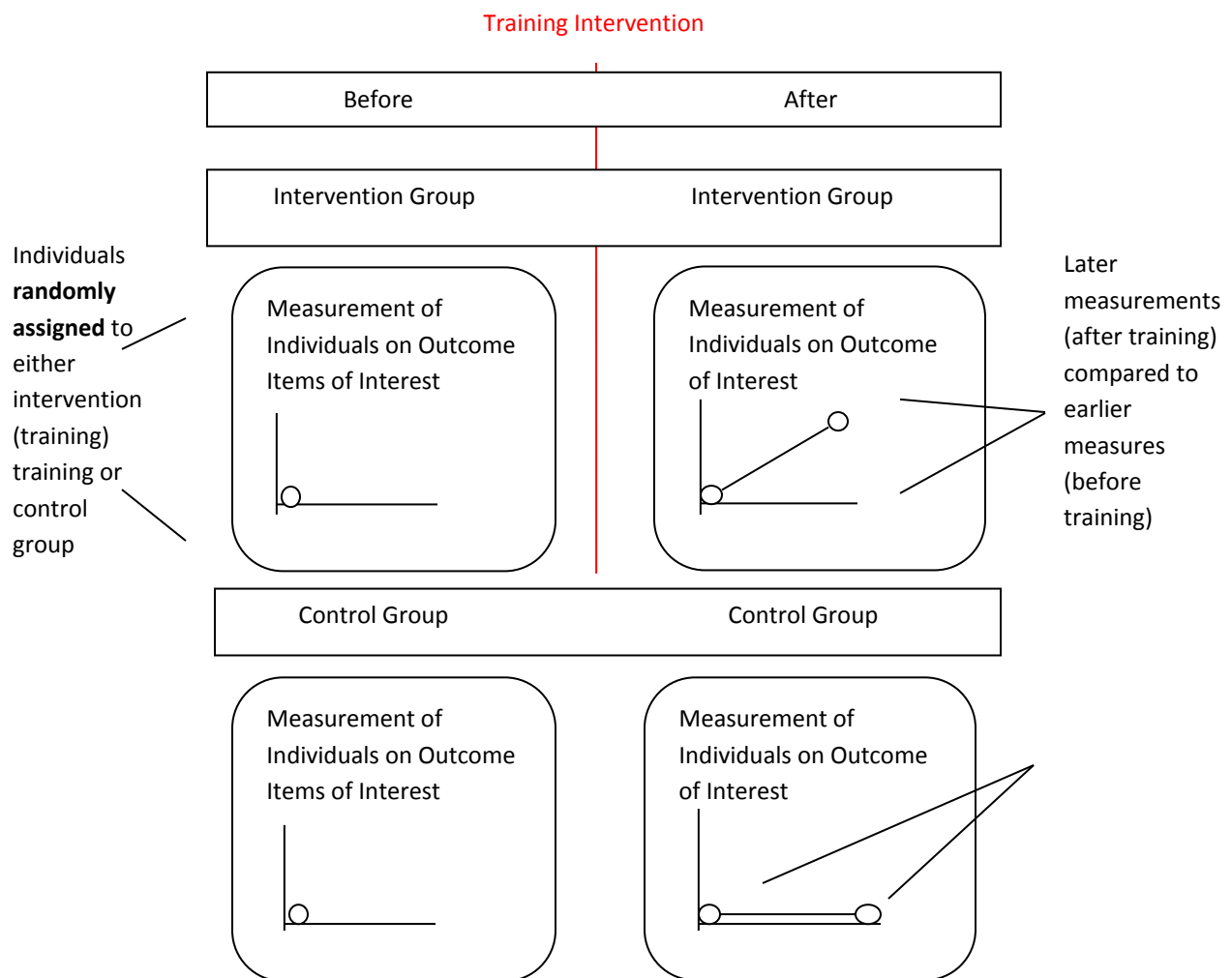


Figure 9

CHAPTER FIVE: THE ASSESSMENT OF TRAINING OUTCOMES

Quasi-Experimental Research Designs – Quasi-experimental approaches are distinguished from true experiments by the lack of random assignment to control and intervention groups. While there are a number of quasi-experimental approaches, the most basic ones with relevance for training evaluation are highlighted below.

Nonequivalent group, post-test only – The nonequivalent, post-test only design consists of administering an outcome measure to two groups or to a program/treatment group and a comparison group. For example, one group of training participants might receive instruction using a new curriculum while the other receives training using the old curriculum. After six weeks, a skills test is administered to both groups to see which curriculum was more effective. A major problem with this design is that the two groups might not be necessarily the same before any instruction takes place and may differ in important ways that influence what skills they are able to assimilate or what progress they are able to make. For instance, if it is found that the participants exposed to the new curriculum perform better, there is no way of determining if they were better prepared or more highly skilled even before the training program began and/or whether other factors are influential to their skill development. A pre-test measure (or means of determining baseline for each group) would need to have been undertaken.

One-group pre-test and post-test – The one-group pre-test- and post-test research design is one of the most frequently used outcome evaluation designs in social science research and in training evaluation. Pre-test measures are taken on a single group of individuals who then later receive the training intervention. After the intervention (training program) has concluded, post-test measures are taken.

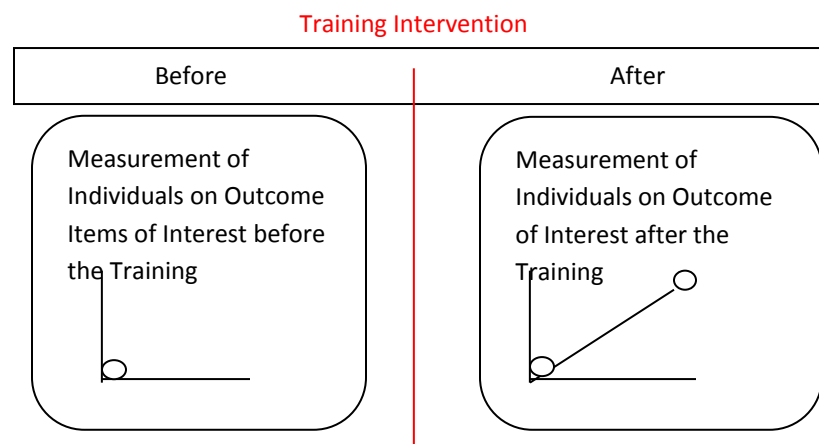


Figure 10

In order to make attributions to the intervention (i.e., training) for any changes in post-test scores compared to pre-test scores, a number of possible confounds or factors must be considered. For example, changes in pre- and post-test scores may be due to *history* in the sense that other events could have happened between the pre-test and post-test that affected individuals' scores. Some of these could have occurred in the work setting (e.g., a new policy might have been implemented that had nothing to do with the training or a change in supervisors took place), while other events could have occurred

CHAPTER FIVE: THE ASSESSMENT OF TRAINING OUTCOMES

outside of the work setting (e.g., a tragic event which ultimately serves as a more powerful motivator to change practice than anything offered in the training program). ⁷¹ In order to rule out these confounding events as possibly affecting post-test scores, you must make the case either that they are implausible in the particular context of the training (e.g., using common sense, theory or experience) or that they are plausible but actually did not have an impact. To make the latter claim, additional research would be needed in order to rule out these alternative explanations for changes in pre- and post-test scores. If you cannot rule out a history threat, then you may report your findings (i.e., changes in measures before and after the training), but also point out the limitations, admitting that you cannot draw confident causal conclusions.

Nonequivalent group, pre-test and post-test – In this design either a naturally occurring group is located or one is constructed, which is similar in as many ways as possible to the group receiving the intervention (i.e., the training program). The only difference you are striving for between the groups is the fact that only one group is receiving the intervention (i.e., the training). For instance, one group could be judges who are as similar as possible to the judges receiving the training but who have not received the training (e.g., same level of experience or previous training, similar caseloads). Or, it could be individuals with the same professional role from different counties within the same state with similar demographic characteristics. Problems arise for constructed matched comparison group designs because (in contrast to true experiments) the comparison group is more likely to be different from the control group. When using this design, care must be taken to ensure that your comparison group is as similar as possible to your intervention group.

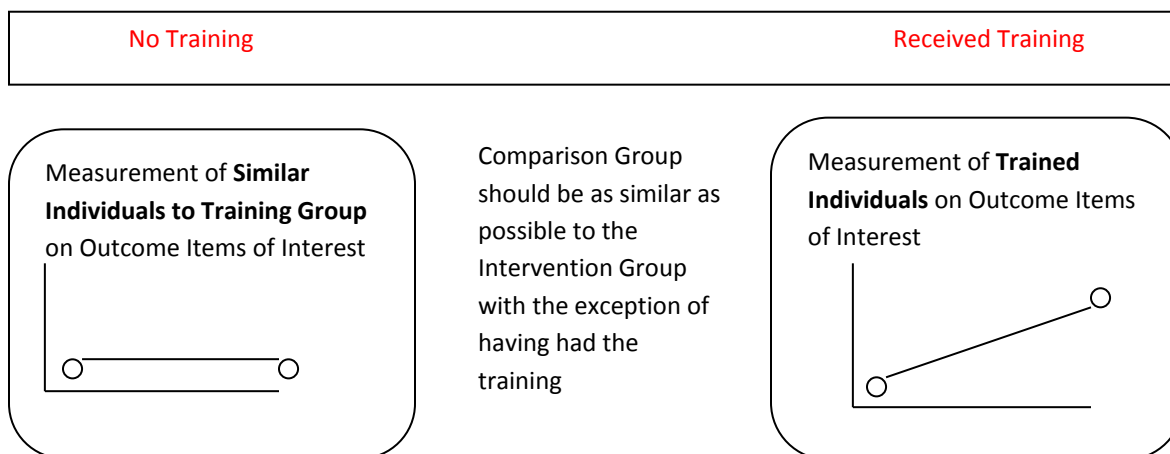


Figure 11

The nonequivalent group, pre-test and post-test design partially eliminates a major limitation of the nonequivalent group, post-test only design. Ideally, at the start of the evaluation study, the researcher

⁷¹ A "history" effect is not the only possible confounding factor but is offered here as an example. A detailed discussion about all of the possible confounding factors with quasi-experimental designs is beyond the scope of this Guide. For more information readers are referred to: Cook, T.D. & Campbell, D.T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin.

CHAPTER FIVE: THE ASSESSMENT OF TRAINING OUTCOMES

empirically assesses differences in the two groups. Therefore, if the researcher finds that one group performs better than the other on the post-test, she can rule out initial differences (if the groups were in fact similar on the pre-test) and normal development (e.g. resulting from practice experiences or previous instruction) as explanations for the differences.

Some problems still might result in this design from participants in the comparison group being incidentally exposed to the treatment condition, being more motivated than participants in the other group, having more resources or organizational support, etc. Additional problems may result from discovering that the two groups do differ on the pre-test measure. If groups differ at the onset of the study, any differences that occur in test scores at the conclusion are difficult to interpret.

Using Historical or Existing Data as a Comparison – When it is not possible to locate (or construct) a group of individuals that is comparable to the group of training program participants, historical data can sometimes serve as a benchmark for comparison. For example, it may not be feasible for a training program implementing court-wide practice change to create two groups of training participants for the purposes of evaluation (i.e., one that receives the training and one that does not). In this case, you might rely on data about performance and outcomes maintained by the court (or available to the court) prior to the changes in practice and compare the two groups on those items to the outcomes observed after the training.

Time series designs – In time series designs, several assessments (or measurements) are obtained from the treatment group (the training participants) as well as from the control group (the individuals who are not exposed to the training). This occurs prior to and after the application of the treatment (i.e., the training). In other words, several pre-test measures are taken as well as several post-test measures - a “time series.” The series of observations before and after can provide rich information about participants’ growth and changes in behavior and practice. Because measures at several points in time prior and subsequent to the training program are likely to provide a more reliable picture of achievement, the time series design is sensitive to trends in performance. Thus, this design, especially if a comparison group of similar individuals is used, provides a strong picture of the outcomes of interest. Nevertheless, although to a lesser degree, limitations and problems of the nonequivalent group, pre-test and post-test design still apply to this design.

In Summary ...

The primary factor directing your training evaluation design is the purpose for the evaluation. Restated, it is critical to consider the utility of any evaluation information. If the program's impact on participant outcomes is a key concern or if multiple programs (instructional strategies, or something else) are being considered and training managers are looking for evidence to assess the relative effectiveness of each to inform decisions about which approach to select, then experimental designs are appropriate and necessary. Nonetheless, resulting information should be augmented by rich descriptions of the training intervention itself and its implementation.

CHAPTER FIVE: THE ASSESSMENT OF TRAINING OUTCOMES

Outcome-oriented training evaluation should be driven by a theory for learning – a theory for how and why the training program will result in change or impact a problem, and what those changes and impacts might be. A theory of learning for your training program (or theory for intervention) not only explains how learning is intended to take place but also why learning and change occurs. These theories provide us with a relevant conceptual framework for interpreting the learning processes and direct our attention to those variables that are crucial in achieving the desired training goals.

In addition to using multiple evaluation methods, evaluators should be careful in collecting the right kinds of information when using experimental or quasi-experimental frameworks. Measures must be aligned with the program's goals or objectives. Additionally, it is often much more powerful to employ multiple measures. Triangulating several lines of evidence or measures in answering specific evaluation questions about program outcomes increases the reliability and credibility of results.

Always consider alternative explanations for any observed differences in outcome measures. If the treatment group (your training group) outperforms the control group, consider a full range of plausible explanations in addition to the claim that the innovative training practice is more effective. Program staff and participants can be very helpful in identifying these alternative explanations and evaluating the plausibility of each. In addition, be upfront about any limitations of the research design you implemented in the outcome evaluation.

SECTION 5.4

Data Collection Methods for Assessing Training Outcomes

Which methods are right for the task? The basic rule here is that the selection of methods follows the selection of focus, not the other way around. Each evaluation question must be examined in relationship to what would constitute evidence for answering it. The following brief descriptions of data collection methods, although by no means exhaustive, can be used as strategies for a variety of circumstances. The list encompasses many of the methods including secondary source or document analysis, observations, interviews, surveys, and focus groups already introduced in previous Chapters of the Guide.

Basic Data Sources for Outcome Measurement

Organizational Records/ Documents Review – Examples of organizational documents that may prove helpful to an outcome analysis include minutes of meetings, policy statements, practice protocols, and case files. These can be treated as data, analyzed for content, and summarized in relation to research questions, including extent of inputs into these materials from the training program's content (i.e., the degree to which trained concepts, tasks, skills and behaviors have been diffused into practice as evidenced by their appearance in the documents studied). Some documents, such as court and agency case files, can also be a rich source for outcome or results measurement. Case files can be examined, for example, for timeframes to permanency, timeliness of court events, etc. Be sure to consider records or documents outside of your own organization as well that might be helpful to obtaining data on your outcome indicators (e.g., school records). If outside organizational records are of value, be sure to plan sufficiently ahead of time for their access. Cooperative or Data Use agreements with other entities may be required and assurances with respect to the confidentiality of information.

Automated Management Information Systems – Court and other partners' management information systems can be a rich source of outcome data and can be analyzed with respect to outcomes of interest to the training (e.g., safety, timeliness, due process and permanency outcomes).

Observations – Some of your outcome indicators may require observation of practice and behavior. Observers can be outsiders or persons who are involved in learning activities. Observers are usually given a list of items that may include extent of participation and personal interaction, nonverbal indicators of interest or inattention, leadership roles, performance levels, and conflict indicators, to name a few. Both qualitative and quantitative data can be collected through an observation process. Ideally a team of observers is used with each observer using the same rating criteria to code an observed event. The goal is to have different trained observers rate the same condition with approximately the same rating, as well as to rate the same condition found in future observation periods with approximately the same rating.

CHAPTER FIVE: THE ASSESSMENT OF TRAINING OUTCOMES

Observations of process and outcomes can be recorded by video or photo documentation. These data are very powerful graphic ways of communicating the nature of a training program's intervention and its outcomes. Observational records of learners' local knowledge and application of behavior changes in the field-setting can also be used to help them reflect on strengths and limitations of knowledge.

Interviews – Interviews and surveys are probably the most widely used method for training program evaluation, including the evaluation of learning and behavior change that may result from training. If the interview questions are standardized, responses can be tabulated numerically to indicate item strength. If questions are open-ended, in-depth unique responses can be generated, which in turn can provide information regarding reasons why training concepts were viewed and applied differently by diverse groups of participants. Focus groups (sometimes referred to as group interviews) can be formed to discuss specific evaluation questions. The purpose of these focus groups is not only to generate judgments using agreed upon criteria, but also to uncover unanticipated outcomes, applications, opportunities, and problems to inform future training efforts.

Surveys – The survey is a standardized form of data collection that incorporates a prepared questionnaire. Many of your outcome indicators may require you to obtain information from the consumers of your training program. For example, your outcome indicator may call for counting the number and percentage of training participants who have implemented a specific strategy learned at the training and changed their behavior. Surveys at the results or impact evaluation level of measurement are often used to evaluate the extent of practice change. They provide a wealth of information that can help in improving training, gauging success levels, and planning for new training programs. Evaluation of unique practices and adaptations of learning, however, is best done through interviews and observations. If you have a large number of individuals to collect detailed information from, it may be best to draw samples of individuals to receive your survey (e.g., seek feedback from every third individual).

Case Studies – In-depth case studies of specific training participants or practices can also be undertaken. Detailed comparisons between individuals who have been part of the training and those who have not are common case studies in the training evaluation context. Case studies are best constructed through repeated interviews over time and often include, in addition to interview or survey self-report, data obtained from persons who know the subjects well (supervisors, mentors, peers). Daily Logs and Activity Journals can also contribute to case study data, with participants documenting their practices over a period of time. Case studies often reveal deterrents to the application of behaviors, as well as ways participants have overcome deterrents to practice.

TOOLS AND RESOURCES – CHAPTER FIVE

- Examples of Outcomes, Indicators, and Data Sources Worksheet
- Key Steps to Implement and Outcomes Measurement Evaluation
- Outcomes, Indicators and Measures - Worksheet
- Tips for Developing Your Outcome Measurement Strategy
- Tips for Identifying Data Sources and Data Collection Methods
- Tips for Identifying Outcomes

CHAPTER FIVE: THE ASSESSMENT OF TRAINING OUTCOMES

CHAPTER FIVE: References and Resources

- Carmona, M.C., Stewart, K., Gottfredson, D.C., Gottfredson, G.D. (1998). *A guide for evaluating prevention effectiveness, CSAP Technical Report*. Rockville, MD: U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Center for Substance Abuse Prevention.
- Cronbach, L. J., & Associates. (1981). *Toward reform of program evaluation: Aims, methods, and institutional arrangements*. San Francisco: Jossey-Bass.
- French, J. F. and Kaufman, N. J. (Eds.) (1981). *Handbook for prevention evaluation: Prevention Evaluation Guidelines*. Washington, D.C.: National Institute on Drug Abuse, Publication No. ADM81-1145.
- Greene, J. C., Caracelli, B. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11, 255-274.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth-generation evaluation*. Newbury Park, CA: Sage.
- Heckman, J.J. (2008). *Econometric Causality*. International Statistical Review, International Statistical Institute, 76(1), pp 1-27.
- Isaac, S. and Michael, W.B. (1983). *Handbook in research and evaluation: A collection of principles, methods, and strategies useful in planning, design, and evaluation studies of education and the behavioral sciences (2nd ed.)*, San Diego, California: EdLTS Publishers.
- Miller, D. C. (1991). *Handbook of research design and social measurement (5th ed.)*. Newbury, Park, California: Sage Publications, Inc.
- Rossi, P. H., & Freeman, H. E. (1982). *Evaluation: A systematic approach*. Beverly Hills, CA: Sage.
- Scriven, M. (1972). Pros and cons about goal-free evaluation. *Evaluation Comment*, 3, 1-5.
- Stufflebeam, D. L. (1971). *Educational evaluation and decision making*. Itasca, IL: F. E. Peacock.
- Tripodi, T, Fellin, P., & Epstein, I. (1971). *Social program evaluation: Guidelines for health education and welfare administration*. Itasca, IL: F. E. Peacock.
- Worden, P. E., & Neumaier, P. (1987). *A source book for program evaluation and accountability*. Fort Collins: Colorado State University, State Cooperative Extension
- W.K. Kellogg Foundation. (1998). *W. K. Kellogg Foundation evaluation handbook*. Battle Creek, Michigan: Collateral Management Company.
<http://www.wkkf.org/documents/wkkf/evaluationhandbook/evalhandbook.pdf>

CHAPTER SIX

Analyzing, Interpreting, and Reporting Training Evaluation Data

Once you have generated data on satisfaction, learning acquisition, and behavior and practice changes and, ultimately, the outcomes you have identified as relevant for your training, you must analyze and interpret those data. In addition, you need to report your evaluation findings in a meaningful way to potentially varied audiences. This Chapter of the Guide covers the steps involved in analyzing and interpreting evaluation data, including providing suggestions for how best to present the results of your training evaluation.

SECTION 6.1

Analyzing and Interpreting Evaluation Data

In this Section, we will cover some of the most common quantitative analysis procedures that are used in training evaluations. Analyzing quantitative and qualitative data is often the topic of advanced research and evaluation methods, and you may need the assistance of experts. While a detailed discussion of analysis is beyond the scope of this Guide - there are certain basics which can help to make sense of reams of data that may be generated from your data collection process. Descriptions of these basics follow.

Whether you have collected qualitative or quantitative data (and ideally you have collected both), your analysis must be logical, thorough, and systematic. Your analysis should also be as simple as possible. For example, many training programs may only need to report descriptive statistics to demonstrate change (i.e., frequencies or percentages), and may have insufficient sample sizes to permit standard tests for significance. For other training evaluations with larger sample sizes and quasi-experimental designs, quantitative data can be analyzed using tests to determine whether there are statistically significant differences between pre- and post-test measures and between training and control groups.

Always Start with your Evaluation Goals and Objectives

When analyzing data (whether from questionnaires, interviews and focus groups, case file reviews or management information systems), always start from a review of your evaluation goals and objectives (i.e., the reason you undertook the evaluation in the first place). This will help you organize your data and focus your analysis. For example, if you wanted to improve your training program by identifying its strengths and weaknesses, you can organize data into program strengths, weaknesses and suggestions to improve the program. If you are conducting an outcomes-based evaluation, you can categorize data according to the indicators for each outcome.

Quantitative Analysis

Quantitative data analysis is helpful in evaluation because it provides quantifiable and easy to understand results. In training evaluation you could have generated “quantifiable” data from a number of data sources, but almost surely you generated quantifiable data from your survey process. These quantitative data can be analyzed in a variety of different ways.

Identify the Type of Data you Have

Because the type of data you have influences the type of analysis you can use, before you begin your data analysis you must determine the level of measurement, or type of data, you have generated in your data

CHAPTER SIX: ANALYZING, INTERPRETING, AND REPORTING TRAINING EVALUATION FINDINGS

collection process. There are four types of data:

- Nominal Data - nominal data are basic classification data (e.g., “male,” or “female”). There is no order associated with nominal data. With nominal data you would assign each category with an arbitrary value (male = 1, female = 0).
- Ordinal Data - ordinal data have a logical order but the differences between the values are not constant (e.g., “small,” “medium,” “large”).
- Interval Data - interval data are continuous and have a logical order. There are standardized differences between the values, but no natural zero (e.g., items expressed on a Likert scale where individuals are asked to rank their choice from 1-5 with 1 being “very dissatisfied” and 5 being “very satisfied”).
- Ratio Data (scale) - ratio data are continuous, ordered, have standardized differences between values, and have a natural zero (e.g., age, length, time).

Once you have identified your type(s) of data, you can begin using some of the quantitative data analysis procedures outlined below. If you have a small sample size, the types of quantitative methods at your disposal are limited. However, there are several procedures you can use to determine what your data are telling you.

Tabulate your Data

- Data tabulation (frequency distributions & percent distributions)
- Descriptive data
- Data disaggregation
- Moderate and advanced analytical methods

The first thing you should do with your data is tabulate your results for the different variables in your data set. Tabulating your data involves creating frequency and percent distributions. This process will give you a comprehensive picture of what your data look like and assist you in identifying patterns. The best ways to do this are by constructing frequency and percent distributions.

A *frequency distribution* is an organized table of the number of individuals or scores located in each category (see the table below). A Frequency distribution table will help you determine if your scores are entered correctly; if scores are high or low; how many scores are in each category; and the spread of scores.

CHAPTER SIX: ANALYZING, INTERPRETING, AND REPORTING TRAINING EVALUATION FINDINGS

Table 6-1 Sample Frequency Table	
Sample Size = 50 completed and returned training surveys	
Role	Judge = 30 Prosecuting Attorney = 20
Jurisdiction	Clark County = 40 Washoe County = 10
Participated in the Workshop on Permanency Hearings Best Practices	Yes = 25 No = 25
Satisfied with Overall Training Program Experience	Very Satisfied = 25 Satisfied = 15 Dissatisfied = 7 Very Dissatisfied = 3

In Table “Title”, you can see that of the training participants who completed and returned a survey, 40 expressed satisfaction with the overall training program experience (with 25 noting that they were “very satisfied”).

A *percent distribution*, on the other hand, displays the proportion of participants who are represented within each category (see the Table “Title” below).

Table 6-2 Sample Percent Distribution Table	
Sample Size = 50 completed and returned training surveys	
Role	Judge = 60% (n=30) Prosecuting Attorney = 40% (n=20)
Jurisdiction	Clark County = 80% (n=40) Washoe County = 20% (n=10)
Participated in the Workshop on Permanency Hearings Best Practices	Yes = 50% (n=25) No = 50% (n=25)
Satisfied with Overall Training Program Experience	Very Satisfied = 50% (n=25) Satisfied = 30% (n=15) Dissatisfied = 14% (n=7) Very Dissatisfied = 6% (n=3)

In Table “Title” above, you can see that 80% (n=40) of the training participants who completed an evaluation survey reported being satisfied with their overall training experience.

Run Descriptive Data

A descriptive refers to calculations that are used to “describe” the data set. Depending on the level of measurement or type of data you have, you may not be able to run descriptive for all of the variables in your dataset. The most common descriptives used are:

- **Mean** – The mean is the numerical average of scores for a particular variable. A meaningful mean can only be calculated from interval or ratio data.

CHAPTER SIX: ANALYZING, INTERPRETING, AND REPORTING TRAINING EVALUATION FINDINGS

- Range – A range express the minimum and maximum values for a variable– the highest and lowest value for a particular variable. A minimum and maximum value can be calculated for all levels of measurement.
- Median – The median is the numerical middle point or score that cuts the distribution in half for a particular variable. A meaningful median can only be calculated from ordinal, interval, and ratio data.
 - Calculate the median by:
 - If the number of scores is odd, the median is the number that splits the distribution
 - If the number of scores is even, calculate the mean of the middle two scores

Mode – The mode is the most common number score or value for a particular variable. A mode can be calculated for all levels of measurement.

Table 6-3 Sample Descriptive Table	
Sample size = 50 completed and returned surveys	
Satisfied with the Overall Training Program Experience	Mean = 3.5 Range = 1 (min) 4 (max) Mode = 3.0 Median = 3.0

Same change as with other tables....From the table above, you can see that the average satisfaction level of the participants who completed an evaluation survey (n=50) was 4.0, with a range from 1 “very dissatisfied,” to 4 “very satisfied.” The most commonly occurring rating (the mode) was a 3 indicating “satisfaction” with the training program.

Disaggregate your Data

Continue to explore your data by disaggregating⁷², or in other words sorting data by different variables and subcategories of interest. Running crosstabs, for example, allows you to stratify data across multiple categories. You may want to break up your data by professional role, by attendance at a particular workshop, by gender or ethnicity, by level of reported experience (such as years on the bench or years of practice), or by jurisdiction.

Using data from the training program evaluation survey example above, we can explore the participants’ role by the jurisdiction they represent. By looking at Table “Title” below, you can clearly see that most of the training participants who returned an evaluation survey were judges (60%; n=30) while just under half of the participants (40%, n=20) were attorneys. The majority of judges (83%) attending the training were from Clark County, while half of the attorneys were from Clark County and half of the attorneys were from Washoe County.

⁷² Disaggregating data is a common exercise of breaking data into subgroups to examine how one group differs from another. Disaggregate and stratify are used interchangeably.

CHAPTER SIX: ANALYZING, INTERPRETING, AND REPORTING TRAINING EVALUATION FINDINGS

Table 6-4 Example table of stratified (or disaggregated) data by region and professional role Sample Size =50 returned and completed surveys [Crosstab: Professional Role by Jurisdiction]				
		Jurisdiction		
		Clark County N=40	Washoe County N=10	Total N=50
Role	Judge N=30	N=25 (83%)	N=5 (17%)	N=30 (60%)
	Attorney N=20	N=15 (75%)	N=5 (25%)	N=20 (40%)
	N= 50 (100)%	N=40 (80%)	N=10 (20%)	N=50 (100%)

You can also disaggregate the data by creating subcategories within a variable. This allows you to take a deeper look at the units that make up that category. From our sample data, 10% of training participants reported that they were dissatisfied with their overall training experience. This subcategory can be explored in more depth to find out more about the make-up of those individuals reporting dissatisfaction with their overall training experience.

Table 6-5 Example table of stratified data by professional role among those who were dissatisfied with training Valid Sample Size = 10 [Participants reporting dissatisfaction with their overall training experience]		
Dissatisfaction with Overall Training Experience	Role	Judge = 0% (n=0) Attorney = 50% (n=10)
	County	Clark = 50% (n=5) Washoe = 50% (n=5)
	Attended Workshop on Permanency Hearing Practice 0% (n=0)	

From this table, you can see that all of the participants who expressed dissatisfaction with their overall training experience were attorneys (n=10). In fact half (50%; n=10 of 20) of all of the attorneys responding to the survey reported that they were dissatisfied. These respondents were equally divided between attorneys from Clark (50%) and Washoe Counties (50%). None of the participants who attended the workshop on best practices for Permanency Hearings reported dissatisfaction with their overall training experience.

Disaggregating data by subcategories within a variable can be helpful in revealing findings that you may not see immediately. For instance, by exploring “dissatisfaction” in more depth in the example above, , you may infer that the training program did not meet the needs of the attorneys that attended, with 50% of the attorneys reporting being dissatisfied with their overall training experience. This result would have been masked if you only reported the average or mean satisfaction level of all training participants, which in this example was a 3.5 on a scale from 1 to 4, indicating satisfaction with the program. It also would have been masked if you had only reported that the majority (80%) of all of the survey respondents reported being satisfied with their training experience. Qualitative responses provided by the attorneys on their survey forms should be explored for any reasons offered as to why they were dissatisfied with their training experience (e.g., perhaps they felt that the curriculum was tailored more to judges and not as relevant to their own professional

CHAPTER SIX: ANALYZING, INTERPRETING, AND REPORTING TRAINING EVALUATION FINDINGS

role). Another finding worth exploring is the fact that none of the individuals who attended the workshop on best practices in Permanency Hearings reported dissatisfaction with their overall training experience, providing some indication that this was a successful workshop.

Basic Analysis of Quantitative Data

- Tabulate the information (add up the number of ratings, rankings, yes's or no's for each evaluation question).
- For ratings and rankings, compute a mean (or average) for each question.
- Also consider reporting the range of answers to an item (e.g., 60% of respondents ranked an item as a "top priority," while 40% of respondents ranked that same item as a "low priority").
- Run cross-tabulations
 - Cross-tabulation is a statistical technique that examines an interdependent relationship between two tables of values but does not identify a causal relationship between the values. For example, a cross-tabulation might show that training participants from a specific stakeholder group were less satisfied with the training content.
- Run statistical tests of significance to compare groups.
 - When you are comparing your training group to a control, or conducting a before or after comparison, the preliminary results are usually summarized into means or scores for each group. Once you've summarized this data, statistical tests of significance should be applied to determine if the observed differences between the two groups' mean scores are real or just a chance difference caused by the natural variation within the measurements (e.g., independent group t-tests or paired t-tests).⁷³

In addition to the basic methods described above there are a variety of more complicated analytical procedures that you can perform with your data. These include:

- **Correlation** A correlation is a statistical calculation which describes the nature of the relationship between two variables (i.e., strong and negative, weak and positive, statistically significant). An important thing to remember when using correlations is that a correlation does not explain causation. A correlation merely indicates that a relationship or pattern exists, but it does not mean that one variable is the cause of the other. For example, you might see a strong positive correlation between participation in the training program and substantive permanency hearing practice. However, the correlation will not tell you if the training program is the reason that permanency hearings have become more substantive. Correlation can imply causation when the data from which the correlate

⁷³ It is beyond the scope of this Guide to provide a primer on statistical techniques. The reader should consult the references at the end of this section for assistance. In addition, consultation with research experts on analytic methods may be required.

CHAPTER SIX: ANALYZING, INTERPRETING, AND REPORTING TRAINING EVALUATION FINDINGS

was computed were obtained by experimental means with appropriate care to avoid confounding and other threats to the internal validity of the experiment.

- **Regression** is an extension of correlation and is used to determine whether one variable is a predictor of another variable. A regression can be used to determine how strong the relationship is between your intervention (i.e., your training) and your outcome variables. More importantly, a regression will tell you whether a variable (e.g., participation in the training) is a statistically significant predictor of the outcome variable (e.g., timely permanency). A variable can have a positive or negative influence, and the strength of the effect can be weak or strong. Like correlations, causation cannot be inferred from regression.
- **Analysis of Variance (ANOVA)** is used to determine whether the difference in means (averages) for two groups is statistically significant. For example, an analysis of variance will help you determine if the mean time to permanency for jurisdictions who received training on permanency hearing best practices is significantly different from jurisdictions who did not receive training on permanency best practices.

Qualitative Analysis

Qualitative data analysis involves the identification, examination, and interpretation of patterns and themes in textual, narrative or “open-ended” data (e.g., respondents' verbal answers in interviews, focus groups, or written commentary on questionnaires), and explores how these patterns and themes help answer the research questions at hand. Qualitative analysis is (NSF, 1997):

- Not guided by universal rules;
- Is a very fluid process that is highly dependent on the evaluator and the context of the study; and
- Likely to change and adapt as the study evolves and the data emerges.

When analyzing qualitative data you must continually reflect back on your evaluation's purpose and goals. The questions to ask throughout the analysis process are (NSF, 1997):

- **What patterns/common themes emerge around specific items in the data?**
 - How do these patterns (or lack thereof) help to shed light on the broader study question(s)?
- **Are there any deviations from these patterns?**
 - If, yes, what factors could explain these atypical responses?
- **What interesting stories emerge from the data?**
 - How can these stories help to shed light on the evaluation questions?

CHAPTER SIX: ANALYZING, INTERPRETING, AND REPORTING TRAINING EVALUATION FINDINGS

- Do any of the patterns/emergent themes suggest that additional data need to be collected?
 - Do any of the evaluation questions need to be revised?
- Do the patterns that emerge support the findings of other corresponding qualitative analyses that have been conducted?

Basic Analysis of Qualitative Data

Qualitative data can produce a wealth of information but not all of it is meaningful. You will need to examine your raw qualitative data to determine what is significant and transform the data into a simplified format that can be understood in the context of your evaluation objectives (Krathwohl, 1998; Miles and Huberman, 1994; NSF, 1997). After you have reduced your qualitative data to its meaningful pieces you will need to group it into meaningful patterns or themes. This process is generally conducted via **content and thematic analysis**.

- **Content Analysis** is carried out by coding the data for certain words or content, identifying their patterns; and interpreting their meanings. This type of coding is done by going through all of the text and labeling words, phrases, and sections of text that relate to your research questions of interest (e.g., “skill acquisition,” “satisfaction,” “behavior change,” etc.). After the data are coded you can sort and examine them for patterns.
- **Thematic Analysis** involves grouping the data into themes that will help to answer your evaluation questions. Once themes have been identified it is useful to group the data into thematic groups so that you can analyze the meaning of the themes and connect them back to the research question(s).

The Basics of Qualitative Data Analysis

- Read through all of the qualitative responses.
- Organize comments into similar categories (e.g., concerns, suggestions, strengths, weaknesses, similar experiences, training program inputs, recommendations, outcome indicators, etc.).
- Label the categories or themes (e.g., concerns, suggestions for improvements, etc.).
- Identify patterns or associations and relationships in the themes (e.g., all people who attended the training who had no previous dependency court experience had similar concerns).

See the Chapter Six Tools and Resources for a summary of the questions to ask throughout the qualitative data analysis process.

CHAPTER SIX: ANALYZING, INTERPRETING, AND REPORTING TRAINING EVALUATION FINDINGS

Interpreting Data

Attempt to put the data obtained in perspective; this could be accomplished by making comparisons. For example, compare results to what you expected or promised and/or original training program goals. You will want to highlight any indications of accomplishing outcomes (especially if you're conducting an outcomes evaluation). In addition, you could provide a description of the training program's experiences, strengths, weaknesses, etc. Consider recommendations to help improve the training program and any conclusions you can draw about program operations or whether training goals were met, etc. Record your conclusions and recommendations in a report document, and associate interpretations to justify your conclusions or recommendations.

In some evaluations, the findings consist of narratives that describe the history of events. In other cases, evaluation findings are descriptions of the way things are without explanations or judgments. In still other cases, evaluations seek to answer questions of cause and effect or the relationships between methods and outcomes. As useful as these findings may be, they do not in themselves help us to judge the worth of a training program's efforts. How do we know that a training program deserves praise? This can occur only when the findings are judged in relationship to several judgment perspectives,⁷⁴ some of which are briefly discussed below.

Standard-Referenced Judgment Perspective – Findings regarding a particular training program's effectiveness can be compared with examples of excellent achievement by using agreed upon criteria established by experts as "state of the art." For instance, such experts could describe excellence as (1) promotion of strategies that are readily applied in the field; (2) recognition of local knowledge and cultural practices; (3) high participation of training participants in designing behavioral change strategies; (4) high inclusion of stakeholders from all system partners; and (5) achievement of targeted behavior and policy changes.

Cohort-Referenced Judgment Perspective – Findings from a training program evaluation can be compared with similar training programs. For this judgment, we ask how a particular training or aspect of a training program would compare to similar training programs elsewhere. External evaluators are most prepared to do this since they may be knowledgeable about training practices and impacts elsewhere. This approach is limited because few places are truly similar in training resources, political and economic contexts, and other conditions that can affect program outcomes. However, a failure to make judgments using external cohort comparisons may result in a false sense of achievement and self-satisfaction. Such training programs, when compared, may be mediocre, outdated, poorly conceptualized, and wasteful while being considered excellent or acceptable by local practitioners and their funders.

⁷⁴ Swanson, Benz & Sofranko (1998).

CHAPTER SIX: ANALYZING, INTERPRETING, AND REPORTING TRAINING EVALUATION FINDINGS

Difficulty-Referenced Judgment Perspective – This perspective takes into consideration the difficulty of what is being attempted when making judgments regarding training program achievements. For example, a training program that is addressing training for systems change with few staff members in the midst of organizational upheaval must be given credit for achievements under these difficult circumstances compared with training programs that have a large staff who are designing and implementing training with relatively rich resources and high motivation for implementing the strategies recommended by the training. Achievement must be judged relative to difficulty and conditions.

Progress-Referenced Judgment Perspective – This perspective on the interpretation of findings gives credit and recognition to progress from past to present. Before-and-after descriptions are essential to making these judgments. Training programs that collect and emphasize "baseline" data usually want to use it to make progress-referenced judgments.

Alternative-Referenced Judgment Perspective – This perspective considers present descriptions of a training program in comparison with what could have been accomplished with the same resources used in alternative ways and places or with different people. This perspective asks, "How else could these resources have been spent with better or different results?" For example, would improved permanency hearing practice that results in shortened timeframes to permanency be more evident today if the training had focused on "front-loading" strategies? What would have been the results if the training program had incorporated all systems stakeholders rather than on just court-based stakeholders? Sometimes answers to these questions can be generated by projecting results based on pilot efforts or by alternative approaches that have already been tried in limited form in other trainings.

These judgment perspectives also are helpful in evaluation of learning by individual participants themselves. Learners can judge their learning in comparison with expert standards, performance by their peers, their external difficulties, their own progress from past to present, and what they would have found more important to have learned or done with their time. Often a combination of these judgments provides a balanced evaluation.

SECTION 6.2

Reporting Training Evaluation Findings

Above all, results need to be presented in a user-friendly or consumable way. If you are leading a training evaluation project, there is nothing more frustrating than finishing your project and finding the comprehensive report you wrote gathering dust on someone's desk. While a comprehensive report might be required and particularly informative for a small number of key stakeholders (e.g., your training organization team, the evaluation team, the funder), a PowerPoint presentation, snapshot or "dashboard" report that summarizes your findings will be more likely to be widely distributed and viewed by many. Sufficient thought and time needs to be spent on the different communication tools that you might need for different audiences. Taking a moment before drafting your report to identify your audience will also assist during the process of writing a report. Maybe your report is only intended for internal stakeholders, or maybe you want to make it available to the public.

Typically, a single report provided to stakeholders neglects the interests of many other participant groups. One solution is to consider the findings, interpretations, and judgments of a training evaluation as constituting a pool from which a variety of reports and styles of reporting can be fashioned to serve specific purposes and different users. Forms of reporting can include formal written reports, written executive summaries, letters to individuals and organizations, exhibits and pictorial displays, bulletins, and public meetings. Stakeholders can then be given the information to which they are entitled in the form that best suits their purposes and best encourages learning.

Sample Training Evaluation Report Outline

There are many possible report formats to consider - a snapshot or "dashboard" report of key findings may be most suitable for certain audiences (i.e., your training participants or system partners), for example, while a comprehensive and detailed report of findings and lessons learned about training program improvement is most suitable for training managers and for funders. Typical sections of a full summary or comprehensive report are noted below.

Title and Opening Pages

The title and opening pages to a comprehensive training evaluation report should provide the following basic information: 1) Name of the training program evaluated; 2) Time-frame of evaluation and date of report; 3) Names and organizations of evaluators; 4) Name of the organizations funding or commissioning the evaluation; 5) Acknowledgements; and 6) Table of contents.

CHAPTER SIX: ANALYZING, INTERPRETING, AND REPORTING TRAINING EVALUATION FINDINGS

Executive Summary

The Executive Summary is an abbreviated version of the most important parts of your evaluation report. The busy reader should come away with an understanding of what the evaluation was about, the main evaluation questions, key findings, and major conclusions and recommendations. Everything in the Executive Summary should be based directly on what is in the report. No new information should be presented in the Executive Summary. The balance of successes and problems documented in the full report should be evident to someone who only reads the Executive Summary. Generally, an Executive Summary should be between two and three pages.

Introduction

The introduction should inform the reader about the context in which the training took place. This would include a summary of the history of curriculum development, the audience for the training, as well as the issues or problems that the curriculum is designed to address. This section of your report should describe the training program (its content and delivery), as well as how it was designed (e.g., whether a needs assessment was used) and identified learning objectives.

The theory of change and logic model for the training program should also be summarized. The reader should know all that is necessary about the context and the problem the training was designed to address. Now the reader wants to know what the training tried to do – laying out the training design and implementation structure helps the reader to understand your evaluation findings and conclusions.

What was the training activity or program about? In this section the evaluator should provide the reader a concise picture of:

- What the training was going to do;
- What the objectives were;
- How it was to be done;
- Where it was to be done;
- Who was going to do it; and
- At what cost (who was funding it)?

Methodology

The credibility of any evaluation's conclusions rests on the quality of the evidence that supports them. This, in turn depends on the appropriateness of the research design and methodology for data collection and analysis

Executive summary

A stand-alone section of two to three pages that should:

- Briefly describe the training program being evaluated (goals, objectives);
- Explain the purpose and objectives of the evaluation, including the audience for the evaluation and the intended uses;
- Describe key aspects of the evaluation approach and methods;
- Summarize principle findings, conclusions, and recommendations.

CHAPTER SIX: ANALYZING, INTERPRETING, AND REPORTING TRAINING EVALUATION FINDINGS

used by the evaluation team. It is important that your report clearly explain the methods used to collect the data in an evaluation.

In this section of your report, the evaluation design (quasi-experimental, mixed methods, etc.) and the methods used to collect data should be presented in summary form. The description should include the unit of analysis, selection of samples (if any), data collection instruments, types of data collected, analytic techniques used, who did it, and when it was done. This can become a rather lengthy description. If so, summarize it and place supplementary information on these points in an appendix. Include questionnaires, observation checklists, descriptions of sampling procedures, data analysis procedures, and other supporting materials in the appendix.

If space permits, a very useful summary chart can be displayed which aligns your evaluation questions with the data type and source(s) used to answer each question.

Limitations of your evaluation method should also be noted. It is important to identify in summary form any data limitations of which your evaluation team is aware. Most training evaluations are constrained by limited time and resources. These influence the overall research design and data collection methods and have an impact on the validity of the findings and conclusions. Other factors such as the timing of the evaluation, database limitations, and accessibility of data may affect the quality of the evidence an evaluation team gathers and the validity of the conclusions reached.

Main Findings

The data from each of your evaluation questions can be summarized from two perspectives. Summarizing topic-by-topic gives information useful for *program improvement*. Summarizing across all topics gives information more useful for *accountability*, such as reporting to stakeholders and funding sources.

Topic-by Topic-Results

- Are suited for instructors and curriculum designers.
- Are most useful for future program improvement.
- Help training coordinators to identify topics which may need improvement.

Across Topics Results

- Give a more succinct picture of the whole program.
- Are suited for reporting to stakeholders and funding sources.
- Lend themselves to simple pie charts for stakeholders or funders who may not be interested in all the details of the first type of training summary.
- Report on Participants' Profile(s)
- Benefits Gained from Attending the Training
- Outcomes at the Individual Level
 - Behavior and Practice Change
- Outcomes at the Organizational Level
 - Impacts on Collaboration
- Training Influences on Outcomes of Interest
 - Training Influences on Outcomes of Interest Identified by Safety, Permanency, Timeliness, Due Process and Well-Being Impacts

CHAPTER SIX: ANALYZING, INTERPRETING, AND REPORTING TRAINING EVALUATION FINDINGS

Conclusion and Recommendations

Interpretations and conclusions drawn from your analysis of the data should be presented in this final section of your report. Conclusions involve judgments and emerge from evaluative reasoning. In the conclusion section, the evaluation team must set forth its deductions about why a training program succeeded or failed to achieve its intended results. Conclusions interpret what your findings mean.

Recommendations and lessons learned regarding what needs to be done to

improve or replicate the training are also relevant for this section of your report. Prior to preparing a set of recommendations, you should review the purpose and goals envisioned by your training program. These purposes should be taken into account as you frame your recommendations for program improvement. Recommendations should:

- Follow directly from your evaluation findings and conclusions
- Be supported by thorough and sound analysis and evaluative reasoning
- Be actionable

Appendices

Your report appendices should include any supporting or explanatory materials that help to inform the reader about the evaluation. Appropriate items for the appendices include:

- Instruments used to collect data
- Data in tables or question-by-question if data are only summarized in Findings section of the report
- Qualitative comments provided by training participants
- The training logic model
- The evaluation plan with specified outcomes, sources for data, data collection methods, who will collect data, etc.
- Reference list and any related resource materials

When preparing your evaluation report:

- Think carefully about the data and findings you want to present. We can often be overwhelmed by data (from survey results for example). If in doubt, put data you consider important but not essential in report appendices.
- Make the evaluation report attractive and easy to read; facilitate this by summarizing the main points and creating a brief presentation.
- Use graphic representations or charts to present findings but be careful not to make these charts too complicated.
- Organize an event such as a staff or team meeting to discuss the results; this could have more impact than the written document.
- Through blogs and wikis, use the evaluation results to generate more discussion and interest in the given subject.

See the Chapter Six Tools and Resources for an additional example of the contents of a comprehensive training evaluation report.

CHAPTER SIX: ANALYZING, INTERPRETING, AND REPORTING TRAINING EVALUATION FINDINGS

TOOLS AND RESOURCES – CHAPTER SIX

- Examples of the Contents of an Evaluation Report
- Questions to Ask Throughout the Qualitative Analysis Process

CHAPTER SIX: ANALYZING, INTERPRETING, AND REPORTING TRAINING EVALUATION FINDINGS

CHAPTER SIX: References and Resources

American Evaluation Association (AEA). <http://www.eval.org/>

Andrews, F.M., Lem, L., Davidson, T.N., O'Malley, P., Rodgers, W.L. (1978). *A Guide for Selecting Statistical Techniques for Analyzing Social Science Data*, Ann Arbor, Michigan: Survey Research Center, Institute for Social Research, University of Michigan.

Boulmetis, J., & Dutwin, P. (2005). *The ABCs of evaluation: Timeless techniques for program and project managers* (2nd ed.). San Francisco, CA: Jossey-Bass

Carmines, E.G., & Zeller, R.A. (1979). Reliability and validity assessment. *Quantitative Applications in the Social Sciences*, 17. Thousand Oaks, CA: Sage Publications, Inc.

Fitzpatrick, J.L. Sanders, J.R., Worthen, B.R. (2004). *Program evaluation: Alternative approaches and practical guidelines*. Third Edition. Allyn and Bacon, Inc.: White Plains, N.Y.

Frierson, H., Hood, S., & Hughes (2002). *The 2002 user-friendly handbook for project evaluation*. National Science Foundation, Directorate for Education and Human Resources, Division of Research, Evaluation, and Communication. NSF 02-057. (http://www.nsf.gov/pubs/2002/nsf02057/nsf02057_1.pdf)

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth-generation evaluation*. Newbury Park, CA: Sage.

Krathwohl, D. R. (1998). *Methods of Educational and Social Science Research: An Integrated Approach* (Second Edition). New York, NY: Longman.

Mertens, D.M. (2005). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. Sage Publications: Thousand Oaks, CA: Sage Publishing Company.

Miles, M.B. & Huberman, A.M. (1994). *Qualitative Data Analysis*. (2nd Ed.). Newbury Park, CA: Sage.

National Science Foundation. (1997). *User Friendly Handbook for Mixed Methods Evaluations*. <http://www.nsf.gov/pubs/1997/nsf97153/>

Rossi, P. H., & Freeman, H. E. (1982). *Evaluation: A systematic approach*. Beverly Hills, CA: Sage.

Spaulding, D.T. (2008). *Program evaluation in practice: Core concepts and examples for discussion and analysis*. San Francisco, CA: Jossey-Bass.

Taylor-Powell, E. and Renner, M. (2003). *Analyzing Qualitative Data*. Madison, WI: University of Wisconsin Extension. <http://learningstore.uwex.edu/pdf/G3658-12.pdf>

United Way of America. (1996). *Measuring Program Outcomes: A Practical Approach*. Alexandria, VA: United Way of America.

Worthen, B.R., & Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidelines*. New York: Longman.