



---

## **Training Evaluation Framework Report**

---

**December 20, 2004**

*Submitted by Cindy Parry, Ph.D., and Jane Berdie, M.S.W.,  
Contractors to CalSWEC for Child Welfare Training Evaluation*

***\*This final draft was submitted to CDSS for review on 12/20/04.\****

# Table of Contents

<b>Executive Summary</b> .....	<b>1</b>
Benefits of Implementing a Framework for Training Evaluation.....	1
Focus of the Framework.....	1
Implementation to Date .....	1
Next Steps.....	2
<b>Introduction</b> .....	<b>3</b>
Overview.....	3
Purpose and Background of the Child Welfare Training Evaluation Strategic Plan .....	3
Benefits of Implementing a Framework for Training Evaluation.....	4
<b>Conceptual Basis for the Common Framework</b> .....	<b>5</b>
Levels of Evaluation .....	5
The Chain of Evidence .....	7
<b>Structure of the Common Framework for Common Core Training</b> .....	<b>10</b>
Overview.....	10
Levels of Evaluation .....	10
<b>Level 1: Tracking Attendance at Training</b> .....	10
<b>Level 2: Course Evaluation</b> .....	11
<b>Level 3: Satisfaction/Opinion</b> .....	12
<b>Level 4: Knowledge</b> .....	12
<b>Level 5: Skills</b> .....	15
<b>Level 6: Transfer</b> .....	17
<b>Level 7: Agency/Client Outcomes</b> .....	18
<b>Bibliography</b> .....	<b>19</b>
<b>Appendices</b> .....	<b>i</b>
Appendix A: Child Maltreatment Identification: A Chain of Evidence Example.....	ii
Appendix B: CDSS Common Framework for Assessing Effectiveness of Training: A Strategic Planning Grid Sample Framework for Common Core .....	vii
Appendix C: Big 5 Curriculum Development and Evaluation Considerations .....	xiv
Appendix D: Teaching for Knowledge & Understanding as a Foundation for Evaluation .....	xix

Appendix E: Writing Curricula at the Skill Level as a Foundation for Embedded Evaluation .....	xx
Appendix F: Bay Area Trainee Satisfaction Form .....	xxiv
Appendix G: Protocol for Building and Administering RTA/County Knowledge Tests Using Examiner and the Macro Training Evaluation Database.....	xxvii
Appendix H: Standardized ID Code Assignment & Demographic Survey .....	xxx
Appendix I: Moving from Teaching to Evaluation—Using Embedded Evaluation to Promote Learning and Provide Feedback .....	xxxiii
Appendix J: Steps in Designing Embedded Evaluations .....	xxxvi
Appendix K: Embedded Evaluation Planning Worksheet for Child Maltreatment Identification at the Skills Level of Learning .....	xl

## **Executive Summary**

Over the past two years the California Macro Evaluation Subcommittee of the Statewide Training and Education Committee (STEC) has engaged in a strategic planning process for training evaluation, resulting in a common framework for training evaluation. The first application of this framework is for Common Core caseworker training. The framework is directly responsive to the California Department of Social Services (CDSS) Program Improvement Plan (PIP) requirement for “assessing the effectiveness of training that is aligned with the federal outcomes.”

### **Benefits of Implementing a Framework for Training Evaluation**

Implementing a multi-level evaluation plan for child welfare training has numerous benefits:

- There will be data about effectiveness of training at multiple levels (a chain of evidence) so that the overall question about the effectiveness of training can be better addressed.
- Data about training effectiveness will be based on rigorous evaluation designs.
- Curriculum writers and trainers will have data focused on specific aspects of training, allowing for targeted revisions of material and methods of delivery.
- Since in California child welfare training is developed and/or delivered by various training entities (RTAs/IUC, CalSWEC, counties), there are many variations of training. Evaluation provides a standardized process for systematic review and evaluation of these different approaches.

### **Focus of the Framework**

The plan addresses assessment at seven levels of evaluation, which together are designed to build a “chain of evidence” regarding training effectiveness. These levels are:

- Level 1 Tracking attendance
- Level 2 Formative evaluation of the course (curriculum content and methods; delivery)
- Level 3 Satisfaction and opinion of the trainees
- Level 4 Knowledge acquisition and understanding of the trainee
- Level 5 Skills acquisition by the trainee (as demonstrated in the classroom)
- Level 6 Transfer of learning by the trainee (use of knowledge and skill on the job)
- Level 7 Agency/client outcomes – degree to which training affects achievement of specific agency goals or client outcomes.

### **Implementation to Date**

The following aspects of the plan are under way:

- Level 1 A system has been designed to gather and transmit this information to CDSS.
- Level 2 The Curriculum Development Oversight Group (CDOG), a subcommittee of STEC, is developing standards and processes for common curriculum in five

priority areas (referred to as the Big 5). These are human development, risk and safety assessment, child maltreatment identification, case planning and management, and placement/permanence. Each of the RTAs/IUC is taking the lead on writing these common curricula.

- Level 3 Each RTA/IUC or county providing training uses an evaluation form for this purpose.
- Level 4 Five core content areas (the Big 5) have been identified as priorities for evaluation at this level. Approximately 250 multiple choice items have been written, reviewed, and researched for evidence based practice in the five priority content areas. Item banking software to manage the test construction, validation, and administration processes has been selected and purchased, and initial training in its use has been conducted.
- Level 5 The Macro Evaluation Subcommittee has selected a priority area for skills evaluation (child maltreatment identification), and has made decisions on initial design considerations. The RTA responsible for development of the child maltreatment identification curriculum is responsible for carrying this evaluation forward with the assistance of CalSWEC.
- Level 6 Currently there are two projects under way at this level of evaluation. In the first, phase one of the evaluation of mentoring programs is nearing completion with responses from 141 caseworkers and 82 supervisors. In the second, the lead agencies responsible for common curriculum in the five priority areas have committed to develop and incorporate transfer of learning activities in the curricula produced.
- Level 7 The above levels are designed to build a “chain of evidence” necessary to provide a foundation for future linking of training to outcomes for children and families. Establishing that training leads to an increase in knowledge and skills that is then reflected in practice is an important part of the groundwork for tying training outcomes to program outcomes that is being laid by the field as a whole.

### **Next Steps**

Over the next six months, the new curricula in the five priority areas and accompanying knowledge tests and transfer of learning activities will be pilot tested. Additionally, the child maltreatment identification skills assessment will be developed and piloted and phase one of the mentoring evaluation will be completed.

During the following six months (July–December 2005) data from knowledge and skills tests will be analyzed, leading to initial validation of assessment instruments and protocols. A process for using assessment findings to review and revise curricula will be developed. Phase two of the mentoring study will be designed to measure the effect of mentoring on the transfer of a specific skill from the classroom to the job.

## **Introduction**

### **Overview**

This report presents the strategic plan for multi-level evaluation of child welfare training in California and describes progress to date. The report covers:

- the need and requirements for child welfare training evaluation in California,
- the structure and processes for development of the strategic plan,
- the rationale for the plan,
- the specific provisions of the plan, including tasks and timeframes, and
- progress to date on implementing the plan.

### **Purpose and Background of the Child Welfare Training Evaluation Strategic Plan**

The purpose of the strategic plan for training evaluation is to develop rigorous methods to assess and report effectiveness of training so that the findings can be used to improve training and training-related activities (such as mentoring and other transfer of learning supports). In doing so, the strategic plan is directly responsive to the California Department of Social Services (CDSS) Program Improvement Plan (PIP), which includes two tasks for training evaluation:

- In consultation with CalSWEC, CDSS will develop a common framework for assessing the effectiveness of training that is aligned with the federal outcomes (Systemic Factor 4, Item 32, Step 1, p. 220).
- CalSWEC and the Regional Training Academies (RTAs/IUC) will utilize the results of the evaluation of the models of mentoring to develop a mentoring component which will be included in the supervisory Common Core Curriculum (Systemic Factor 4, Item 32, Step 4, p. 222).

The strategic plan is a common framework for evaluation that addresses multiple levels of evaluation and for each, the decisions/actions, the needed resources, and the timeframes for implementation. It is a framework that can be applied to evaluation of any human services training, although the work to date specifically addresses Common Core Training. The evaluation plan for Common Core Training coincides with a statewide initiative for revision of five major training content areas, which in turn address practice improvement needs highlighted by the Child and Family Service Review (CFSR) and the PIP.

Within the framework, there are several specific projects already under way or proposed for the near future. These include:

- A system for tracking and reporting all caseworker completion of Core Training
- Quality assurance for the development and delivery of revised Core Training in five content areas.
- Development and utilization of knowledge testing for five content areas of Core Training.
- Development and utilization of skills assessment in one area of content of Core Training.

- Assessment of the effectiveness of mentoring as a method to increase transfer of learning for participants in Core Training. The findings from this evaluation will inform the development of the supervisory Common Core Training.
- Analysis of data at multiple levels of evaluation in order to build a “chain of evidence” about training effectiveness in order to answer broad questions about the impact of training on achievement of agency and client outcomes.

The framework for training evaluation was developed by the Macro Evaluation Subcommittee of California’s Statewide Training and Education Committee (STEC). The Macro Evaluation Subcommittee consists of representatives of the five Regional Training Academies/Inter-University Consortium (RTA/IUC), CalSWEC, several county agency training staff, and CDSS. The Macro Evaluation Subcommittee began meeting in June 2002. With the assistance of CalSWEC consultants, the Macro Evaluation Subcommittee first developed parameters for establishing a strategic plan for evaluation of child welfare training and then began addressing the attendant issues. In the 2 ½ years since its inception, the Macro Evaluation Subcommittee has developed the Common Framework Strategic Planning Grid for Core Training Evaluation, has reviewed the status of current evaluation at all seven levels, and has begun or continued implementation of evaluation initiatives in six of the seven levels.

### **Benefits of Implementing a Framework for Training Evaluation**

Implementing a multi-level evaluation plan for child welfare training has numerous benefits:

- There will be data about effectiveness of training at multiple levels (a chain of evidence) so that the overall question about the effectiveness of training can be better addressed.
- Data about training effectiveness will be based on rigorous evaluation designs.
- Curriculum writers and trainers will have data focused on specific aspects of training, allowing for targeted revisions of material and methods of delivery.
- Since in California child welfare training is developed and/or delivered by various training entities (RTAs/IUC, CalSWEC, counties), there are many variations of training. Evaluation provides a standardized process for systematic review and evaluation of these different approaches.

## **Conceptual Basis for the Common Framework**

### **Levels of Evaluation**

Training is traditionally formally evaluated primarily by assessing trainee reactions, i.e., their satisfaction with the training and their opinions about its usability on the job. Informally, it is often evaluated by the trainers and sometimes by an advisory group who view written material and delivery to assess the relevance of content and the degree to which the methods of delivery and content hold the interest of the trainees. Occasionally knowledge is tested using a paper/pencil test.

A more rigorous approach to training evaluation is to identify a continuum of levels of evaluation, determine which are most useful, and design procedures accordingly (Kirkpatrick, 1959; Parry & Berdie, 1999; Parry, Berdie, & Johnson, 2004). The American Humane Association (AHA) offers one model for identifying levels of evaluation. In this model (depicted in the graph below), the levels of evaluation are shown as closer to or farther from the actual training events in order to represent the extent to which factors other than training likely affect the cause-effect relationship between the training and the evaluation findings. For example, when a course curriculum and delivery are evaluated (a formative evaluation), usually very little “interference” is present. However, subsequent levels of evaluation are increasingly affected by intervening variables. Findings about knowledge acquisition and understanding may be affected by trainee learning differences and education levels. Trainee performance on skill assessments may be affected by opportunities for practice before testing, and the effect of training on client outcomes may be impacted by agency priorities and caseload size. The issue of the impact of other events is critical in the decision-making about which levels of evaluation to conduct and the design of the evaluations.

The AHA levels of evaluation are as follows:



The first level, the **Course** Level, includes the evaluation of the training itself: content, structure, methods, materials, and delivery. It may also include evaluation of the adequacy of the outcome measurement tools to be used. Course level evaluation is conducted to guide revisions and refinements to the training in order to maximize its quality and relevance and the attainment of desired trainee competencies. Thus, feedback is usually detailed, descriptive, and narrative in nature.

The second level, **Satisfaction**, measures the trainees' feelings about the trainer, the quality of material presented, the methods of presentation, and environment

(e.g., room temperature).

Level 3, **Opinion**, refers to the trainees' attitudes toward utilization of the training (e.g., their perceptions of its relevance, the new material's fit with their prior belief system, openness to change), as well as their perceptions of their own learning. It goes beyond simply a reaction to the course presentation and involves a judgment regarding the training's value. This level often is measured by questions on a post-training questionnaire or as part of a "happiness sheet" which ask the trainee to make judgments about how much he or she has learned or about the information's value on the job. Like the level above, this measure is self-report and provides no objective data about learning (Johnson & Kusmierke, 1987; Pecora, Delewski, Booth, Haapala, & Kinney, 1985).

Level 4, **Knowledge Acquisition**, refers to such activities as learning and recalling terms, definitions, and facts and is most often measured by a paper and pencil, short answer (e.g., multiple choice) test.

The next level, **Knowledge Comprehension**, includes such activities as understanding concepts and relationships, recognizing examples in practice and problem solving. This level can be measured by a paper and pencil test, often involving case vignettes.

Level 6, ***Skill Demonstration***, refers to using what is learned to perform a new task within the relatively controlled environment of the training course. It requires the trainee to apply learned material in new and concrete situations. This level of evaluation is often “embedded” in the classroom experience, providing both opportunities for practice and feedback and evaluation data (McCowan & McCowan, 1999).

Level 7, the ***Skill Transfer*** level, focuses on evaluating the trainees’ performance on the job. This level requires the trainee to apply new knowledge and skills in situations occurring outside the classroom. Measures that have been used at this level include Participant Action Plans, case record reviews, and observation.

The last three levels in the model, ***Agency Impact, Client Outcomes, and Community Impacts***, go beyond the level of the individual trainee to address the impact of training on child welfare outcomes. Outcomes addressed at these levels might include, for example, the impact of training in substance abuse issues, on patterns of services utilized, or interagency cooperation in case management and referral. Cost-benefit analyses might also be conducted at agency, client, or community levels. At these levels, training is typically only one of a number of factors influencing outcomes. Evaluation should not be expected to unequivocally establish that training, and training alone, is responsible for changes observed. However, training may well play a role in better client, agency or community outcomes. Well-designed and implemented training evaluation can help to establish a “chain of evidence” for that role.

For the purpose of the Common Framework Strategic Planning Grid, some of the levels in this model have been collapsed. For instance “knowledge acquisition” and “knowledge understanding” are called “knowledge” and the design of the knowledge testing captures both levels. A level has been added in the beginning: “tracking” enables CDSS to assess the degree to which all new staff receives Core training in the mandatory timeframe. (See below for a list of the levels in the framework.)

### **The Chain of Evidence**

The chain of evidence refers to establishing a linkage between training and desired outcomes for the participant, the agency, and the client such that a reasonable person would agree that training played a part in producing the desired outcome. In child welfare training, it is often impossible to do the types of studies that would establish a direct cause and effect relationship between training and a change in the learner’s behavior or a change in client behavior, since these studies would involve random assignment. In many cases, ethical concerns would prevent withholding or delaying training (or even a new version of training) from a control group.

To definitively say that training was responsible for an outcome, one would need to compare two groups of practitioners where the only differences between groups was that one received

training and one did not. Random assignment to a training group and a control group is the only recognized way to fully control for all other possible ways trainees could differ besides training that might explain the outcome. For example, in a study designed to see if an improved basic “core” training reduces turnover, many factors in addition to training could affect the outcome. Pay scale in the county, relationships with supervisors and co-workers, a traumatic outcome on a case, or any of a host of personal factors might impact the effectiveness of new trainees. With random assignment, these factors (and any others we didn’t anticipate) are assumed to be controlled, since they would not be expected to occur more often in one group than the other.

Other types of quasi-experimental designs are possible and much more common in applied human services settings. These designs try to match participants on relevant factors besides training or identify a naturally occurring comparison group as similar as possible to the training group. For example, in the turnover study outlined above, we might take several different measures to control for outside factors. We might match participants by pay scale, or we might attempt to control for the supervisory relationship by having trainees fill out a questionnaire on their supervisor and matching those with like scores. It is almost impossible, however, to anticipate and control for all the possibilities and to match the groups on all of the relevant factors.

When we are faced with a situation where quasi-experimental designs are the best alternative, it strengthens our argument that training plays a part in producing positive outcomes if we can show a progression of changes from training through transfer and outcomes for the agency and client. In building a chain of evidence for this example, we might start with theory, pre-existing data (e.g., from exit interviews) and common sense that suggests that having more skill and feeling more confident and effective in doing casework increases a worker’s desire to stay on the job. If we can then establish that caseworkers saw the training as relevant to their work, learned new knowledge and skills on the job, used these skills on the job, and had a greater sense of self-efficacy after training, we have begun to make a logical case that training played a part in reducing turnover. From that point, quasi-experimental designs can be used to complete the linkage. For example, level of skill and efficacy could be one of the predictors in a larger study of what reduces turnover, with the idea that more skilled people will be less likely to leave (other factors being equal).

To achieve a chain of evidence about training effectiveness, it is useful to develop a structured approach to conducting evaluation at multiple sequenced levels (lower levels being those most closely associated with training events). Since higher levels build upon lower levels, it is also necessary to consider whether or not a particular evaluation should collect information at levels lower than the level of primary interest. For example, if the primary focus of a training evaluation was on whether or not a particular skill (e.g., sex abuse interviewing) was used on the job, the evaluation would need to be designed to collect Level 7, Skill Transfer, data. If the evaluation showed that almost all trainees used the new techniques competently, that level of information alone would be sufficient. If, as often happens, the results did not show that

trainees were demonstrating the desired behavior, then the question of, “Why not?” becomes relevant. In order to answer that question, it becomes necessary to step back through the levels and ask: “Did the trainees meet the training objectives and acquire the knowledge and skills in the classroom?” If the answer is *no*, then trainee satisfaction and opinion data may be needed to shed light on the problem. Perhaps the training was not delivered well, or the trainees did not see its relevance or were not open to changing old behaviors. For an example of how the chain of evidence addresses a key child welfare topic (child maltreatment identification), see *Appendix A*.

For these reasons, the Macro Evaluation Subcommittee has decided to include multiple levels of evaluation in the framework.

## **Structure of the Common Framework for Common Core Training**

### **Overview**

The structure of the framework uses seven levels of evaluation as the major components. These levels are:

- Level 1 Tracking attendance
- Level 2 Formative evaluation of the course (curriculum content and methods; delivery)
- Level 3 Satisfaction and opinion of the trainees
- Level 4 Knowledge acquisition and understanding of the trainee
- Level 5 Skills acquisition by the trainee (as demonstrated in the classroom)
- Level 6 Transfer of learning by the trainee (use of knowledge and skill on the job)
- Level 7 Agency/client outcomes – degree to which training affects achievement of specific agency goals or client outcomes.

For each level, the following information is provided:

- The scope, i.e., how much of Common Core Training is being evaluated at this level
- Description of the level, including what is addressed in the evaluation and the tasks to carry out the evaluation
- Decisions – what decisions have been made and what decisions are pending that affect design of the evaluation(s) at this level
- Resources, i.e., what resources are needed to implement evaluation at this level
- Timeframes for the various tasks.

This information is summarized in *Appendix B: CDSS Common Framework for Assessing Effectiveness of Training: A Strategic Planning Grid*, and described in more detail below.

## **Levels of Evaluation**

### **Level 1: Tracking Attendance at Training**

#### ***Scope and Description***

Although this level is not included in the AHA model of levels of evaluation for training, it is an important precursor to a system of training evaluation. A system for tracking attendance at training is needed to ensure that new caseworkers are being exposed to training on all of the competencies needed to do their jobs, and that that occurs consistently across the state. The California Department of Social Services, through CalSWEC and the Statewide Training and Education Committee (STEC), has recommended a Common Core Curriculum that includes sets of competencies, learning objectives, and content resources. At this level, names of individuals completing Common Core will be tracked by the Regional Training Academies (RTAs/IUC) and reported to the counties semi-annually. The counties will provide the State with aggregate numbers of new hires and numbers of workers completing Common Core for

the year as part of their annual training plan. STEC recommends that new hires be required to complete Core training within 12 months.

### ***Resources***

Databases will be required for capturing information for tracking by the RTAs/IUC, counties and the State. RTAs have the capacity to provide this information to counties and already do, however, there may be additional databases and transmittal procedure needs identified for county submissions to the state. These entities will also need to commit personnel time to maintain the databases and monitor submissions for timeliness and quality (both locally and centrally). Protocols and training for individuals involved in maintaining and submitting data also will be developed or maintained by these entities. Methods for tracking data can vary, ranging from paper submissions to an electronic submission such as an Excel spread sheet.

### ***Timeframes***

This level is currently in place and will be used for reporting tracking data with the pilots of the new training (*see description below*), which begin in spring 2005.

## **Level 2: Course Evaluation**

### ***Scope and Description***

The Macro Evaluation Committee of STEC has identified five areas of the Common Core as priorities for some curriculum standardization and for evaluation. These are human development, risk and safety assessment, child maltreatment identification, case planning and management, and placement/permanence (hereafter called “The Big Five”). Within these five areas, additional curriculum development is under way. Under the direction of the Content Development Oversight Group (CDOG), a subcommittee of STEC, each of the Big 5 is a responsibility of one or more of the RTAs or CalSWEC. In each case, the lead entity will review and update existing competencies, learning objectives, and curricula content in order to provide both a high quality training experience for workers and a consistent basis for higher levels of evaluation. The RTAs/IUC and counties have committed to teach this common content, as part of their Core Curriculum. They may also add additional content in these areas if they desire, but the common content must be taught. At the course level, activities are focusing on the further development of these curricula, guided by a process of formative evaluation and quality assurance. All of the RTAs/IUC and counties that provide training will use evaluation data to improve their delivery of Common Core curricula and engage in systematic statewide updates of content and methods.

### ***Resources***

Each lead entity needs to ensure personnel and/or consultants for forming a workgroup, reviewing and revising competencies and learning objectives, reviewing literature as needed, developing new curricula as needed, adhering to CDOG’s decisions and protocols regarding quality assurance and a curriculum format, and working with the training evaluators around

specific knowledge items (and for the RTA taking the lead on Child Maltreatment Identification, also the skills evaluation).

Each lead entity will also need to participate in CDOG meetings to guide and track progress.

### ***Timeframes***

Specific protocols/formats for constructing curricula, reviewing curricula, and observing delivery will be developed by CDOG prior to March 2005 and used in the pilot process. Piloting of draft curricula will begin in March 2005 with a goal of finalizing curricula in the first quarter of FY 05/06.

Additional quality assurance procedures for updating/modifying curricula will also be developed by CDOG during FY 2005-2006. Guidelines include “Big 5 Curriculum Development Evaluation Considerations (*Appendix C*), Teaching for Knowledge and Understanding as a Foundation for Evaluation (*Appendix D*), and Writing Curricula at the Skill Level as a Foundation for Embedded Evaluation (*Appendix E*).

## **Level 3: Satisfaction/Opinion**

### ***Scope and Description***

The RTAs/IUC currently use forms to collect participant feedback on the quality of training. This will continue. No standard form will be required due to local constraints related to University requirements of the RTAs/IUC. Those that wish to may use a standard form that the Bay Area Academy and CalSWEC have developed (*Appendix F*). If an RTA/IUC or county desires to link satisfaction data with the outcomes of knowledge and skill evaluations, they may do so by including the same personal identifier on the satisfaction form. This is not required as part of the framework.

### ***Resources***

No additional resource needs are anticipated for this level of evaluation.

### ***Timeframes***

These evaluations are in place and currently being conducted.

## **Level 4: Knowledge**

### ***Scope and Description***

Knowledge related to key competencies in the Big 5 will be evaluated using multiple choice knowledge tests pre and post training. Feedback provided at this evaluation level will be used

for course improvement and demonstrating knowledge competency of workers in aggregate. Individuals' test results will not be reported to the State or shared with supervisory personnel.<sup>1</sup> The RTAs/IUC will conduct evaluation according to agreed upon protocols and send data to CalSWEC to validate items. CalSWEC will validate the items and update an item bank based on the data.

CalSWEC has developed and will maintain an item bank from which tests will be constructed for these five areas. To date, approximately 250 items have been developed for the five areas. These items have undergone extensive editorial review by the RTAs/IUC, counties, CalSWEC, and consultants with child welfare and test construction expertise. Research or policy bases have been established for almost all of the item content (with the exception of some that seem based on conventional wisdom). The next steps in item validation will be the collection of data from training participants on which a statistical item analysis will be conducted. Items may be added and validated on an ongoing basis as curricula are updated or new methods of training are implemented. Item banking software has been reviewed and a program called EXAMINER has been selected and purchased. RTA and county representatives have received initial training on its use and will receive further training.

For the Big 5, essential information presented in training will be standard (see discussion under level 2). The items and the literature reviews done to establish an evidence base for their content will be used as one source to inform the curriculum development process. The Macro Evaluation Committee has recommended to STEC that as common Big 5 content is developed by each lead agency, that agency will identify a common item set that must be included in the knowledge test for that area. Each RTA and/or county will also have the opportunity to identify new items to be developed. CalSWEC will work with them to develop these items. CalSWEC will take the responsibility of constructing test templates from these common items.

Additionally, RTAs/IUC and counties will have the opportunity to use other items from the item bank as they desire to test other content included in their training modules that is not part of the required common content. They also will be provided copies of the EXAMINER software and may choose to construct their own knowledge items and item banks for any of their courses they wish to evaluate.

A protocol has been agreed to by the Macro Evaluation Committee and recommended to STEC that spells out procedures for test construction and administration (*see Appendix G*). The protocol calls for use of pre and post tests of 25 to 30 items each. Recommended time allowed for each test is 45 minutes for pre and 30 minutes for post. The same test form will be used for pre and post testing. Pre and post tests will be conducted for the Big 5 content areas (where training for a content area is longer than 1 day) until items are validated. Post tests only will be conducted for Big 5 content areas where training for a content area is 1 day or less.

---

<sup>1</sup> The IUC has and will continue to provide knowledge assessment data to trainees' supervisors.

Once items are validated, this decision will be revisited. At that time, the pre-test may be eliminated except for a random sample of training classes. Routine pre-testing may not be needed if: data show that trainees consistently don't know the material prior to training (thus making continued verification of this fact unnecessary), and posttest scores continue to fall in an acceptable range for indicating mastery of the material. Data will be collected from each participant but reported only in aggregate. A confidential ID code will be used to link pre and post tests for analysis and to link test data to demographic data (*see Appendix H for a copy of the Identification Code Assignment and Demographic Survey form*).

Participants will be informed of the purpose of the evaluation, confidentiality procedures and how the results will be reported and used. Trainers will have written instructions and/or training in how to administer and debrief evaluations and monitor the ID process. For security of the item bank items, participants will turn in their tests before leaving the classroom to avoid a loss of item validity if they are circulated.

Following the pilot phase, RTAs/IUC and counties will have access to statewide aggregate data and data from their own trainings. They will use the results to determine the extent to which knowledge acquisition is occurring. During the pilot and item validation, results will not be used to evaluate training effectiveness or trainee learning, as they might be misleading. A timeline for phases of the evaluation process will be developed and shared with counties to clarify what results will be available to them and how they may be used appropriately.

### ***Decisions Pending***

Decisions regarding what type of data entry (manual, scanned locally, scanned centrally) will be most accurate, efficient, and cost effective are pending.

### ***Resources***

A number of costs are associated with this level of evaluation. At the local level, RTAs/IUC and counties have expended staff time to review items for the item bank, and attend item bank training. Staff time will also be needed to prepare and distribute paper tests, and to manage storage and transmission of data. CalSWEC is exploring several options for scoring tests and transmitting data; including the option of purchasing scanners for the RTAs and counties involved in testing, and scanning data centrally at CalSWEC. Regardless of the option chosen, RTAs/IUC and counties will need to allocate staff time for either data entry or scanning of forms and quality assurance. There are also costs associated with copying and mailing of paper forms, and trainer time for learning test administration and transmission procedures.

There are also a number of resource issues assumed by CalSWEC, in addition to the potential costs of purchasing a scanner or scanners. CalSWEC has already purchased the EXAMINER item banking software. Other resources expended have been staff time and associated costs for managing the process, item review, software training, and consultant services for item development and review and software selection. Additional resource needs are anticipated in the areas of statistical validation and scaling of items, data analysis and reporting, item bank

maintenance, distribution of test templates, further item bank training and technical assistance, and coordination of the data submissions.

### ***Timeframes***

Further training on EXAMINER will be provided by CalSWEC and the evaluation consultants just prior to the new curriculum pilots in March 2005. A final version of the protocol for data entry will be shared with RTAs/IUC at this training.

Knowledge testing in the five priority areas will begin in March 2005 concurrent with the pilots of the new common curricula under development by CDOG. Annotated item bank items were provided to the leads for curriculum development from CalSWEC during October, 2004, with the exception of risk and safety assessment which has been delayed somewhat while the effects of a new dual track system of assessment and investigation on training content have been assessed.

## **Level 5: Skills**

### ***Scope and Description***

One trial area, Child Maltreatment Identification, will be evaluated at the skill level. An *embedded evaluation* will be developed which requires that, for this curriculum area, both the content and method of delivery will be standard. Embedded evaluation builds on existing exercises or designs new tasks that can be used as both instructional and evaluation opportunities. This linkage enhances trainee learning and provides relevant feedback to trainers for course improvement, while also providing important data on trainees' acquisition of skills.

Embedded skill demonstration tasks in the classroom are promising for two additional reasons. First, skill level evaluation tasks are time consuming and logistically difficult. Within the training day a performance task that wasn't integrated with instruction would take too much time away from an already tight schedule. Designing classroom exercises that both teach skills and evaluate the training promotes efficiency. It also enhances both trainee learning and the quality of the evaluation data. Second, using embedded performance tasks during training provides a baseline for linking training performance with transfer activities. One necessary prerequisite to transfer of learning to the job is initially having learned the skill. Embedded evaluation can help to document to what extent that learning is taking place in the classroom and to what extent transfer could reasonably be expected to take place even under optimal conditions in the field.

All RTAs/IUC and counties that deliver training will integrate the curriculum and embedded evaluation under development for this skill into existing Core training. The Macro Evaluation Subcommittee has agreed that the evaluation will use case scenarios and slides and will focus on identification of whether physical abuse as defined by California Welfare & Institutions (W & I) Code has occurred.

This evaluation, like evaluation described in level 4, will be used for course improvement and demonstrating competency of workers in aggregate (not individuals). Since there is little likelihood that the majority of participants will have this skill prior to training, the evaluation will be conducted at the end of training only. Pre-testing of skills using performance tasks is usually impractical since it is very time consuming, as well as technically difficult to equate for task difficulty, and therefore costly.

There are a number of steps in designing embedded evaluation (*see Appendices I and J*) and related decisions regarding the skills assessment protocol (*see Appendix K*). The participant ID procedures and information from the demographics form (discussed under level 4 above) will also be used for skill level evaluation. Data will be collected from each participant but reported only in aggregate. Participants will be informed of the purpose of the evaluation, confidentiality procedures and how the results will be reported and used. Trainers will have written instructions and/or training in how to administer and debrief evaluations, and participants will turn in any evaluation forms before the trainer processes the evaluation exercise and will not take any evaluation materials out of the classroom.

As in level 4, RTAs/IUC and counties will have access to statewide aggregate data and data from their own trainings, and will use results to determine the extent to which skill acquisition is occurring.

### ***Decisions Pending***

There are still a number of decisions pending with respect to evaluation of skill at level five. These are dependent on the final recommendations made by the lead agency developing curriculum regarding evaluation scope and design.

- Whether or not to evaluate competencies related to neglect is under consideration by CDOG
- Scoring related decisions - after the item content is finalized, it will be necessary to develop a scoring key and procedures. Issues to consider include the need to develop a minimum competency standard against which to judge performance. In a post test only format, performance is judged against a desired standard, rather than a pre-test performance level. Content experts will work with the consultants to develop a scoring key for the items, a standard for individual performance and a standard for judging the effectiveness of training. Although individual scores will not be shared, in order to determine if a desired percentage of people overall have met the standard, it is necessary to know how many individuals have met the standard.
- As with level 4, decisions regarding the most cost effective way to transmit evaluation data to CalSWEC for analysis.

### ***Resources***

Resources needed at this level include personnel time from the RTAs/IUC, and counties for participation in CDOG curriculum development activities, trainer/subject matter expert time for

consulting on evaluation design and scoring rubrics, and trainer time for learning administration and debriefing of the evaluation. There is also a need for CalSWEC staff and consultant time for evaluation design, analysis and reporting to the RTAs/IUC, counties, and state.

### ***Timeframes***

The lead RTA working on the curriculum revisions will work with CalSWEC and the evaluation consultants to develop the skills evaluation to be ready for piloting along with the curriculum in March 2005.

## **Level 6: Transfer**

### ***Scope and Description***

There are two main activities included under Level 6. One is planned as part of the curriculum development process described in Level 2. The first is that suggested transfer of learning activities will be developed for the Common Core as part of the curriculum development process for the five priority areas.

The second is an evaluation of the role of mentoring programs in transfer of learning. This evaluation is designed in two phases. Two RTAs have participated in phase one of this project which assesses the extent to which the provision of mentoring services:

- increase perceived transfer (by workers and their supervisors) of Core knowledge and skills,
- increase worker satisfaction with the job and feelings of efficacy, and
- contribute to improved relationships with supervisors.

Both mentored workers and new caseworkers who do not receive mentor services are being asked to rate their skills, the supervisory support they receive, and their job comfort and satisfaction at the beginning and end of a six month mentoring period. Their supervisors are also being asked to rate their worker's skills and the supervision they provide. If mentoring is effective, the evaluation should show a larger skill gain for the mentored workers than the workers who are not mentored.

To date, pre-test data has been received from 141 caseworkers and 82 supervisors. Data collection will continue until sufficient posttest and comparison group data are available to complete the analysis. After phase one is completed, one RTA will continue with phase two of the evaluation which will focus in depth on one skill and assess the skill in the classroom and on the job.

### ***Resources***

Resource requirements at this level of evaluation locally include training coordinator and mentor time to: participate in planning the evaluation, review the design and instrumentation,

complete and track completion of data collection instruments, enter data, and participate in project meetings. There is also a significant amount of consultant time required to design the evaluation, develop the evaluation instruments, develop databases and enter data, conduct analyses, and write reports.

### ***Timeframes***

Phase one of the mentoring evaluation began in the fall of 2003 and is anticipated to conclude in April 2005. The timeframe for the development of Big 5 transfer of learning activities will be decided by CDOG. Basic transfer of learning (TOL) materials will be created concurrent with Big 5 content development. After piloting, these TOL materials will be evaluated and refined in subsequent years.

## **Level 7: Agency/Client Outcomes**

### ***Scope and Description***

Using the entire framework, California can begin to build the 'Chain of evidence' necessary to evaluate the impact of training on outcomes. Linking the outcomes of training with program outcomes is a complex process that requires the careful assessment of multiple competing explanations for any change that is observed. Building the supporting pieces necessary to show that training has an effect on worker practice is a first step in linking worker training to better outcomes for children and families. California has chosen to focus on this necessary groundwork in this framework document to build a firm foundation for future efforts to evaluate at this level.

## **Bibliography**

- Johnson, J., & Kusmierek, L. (1987). The status of evaluation research in communication training programs. *Journal of Applied Communication Research*, 15(1-2), 144-159.
- Kirkpatrick, D. (1959). Techniques for evaluating training programs. *Journal of the American Society of Training Directors*, 13(3-9), 21-26
- McCowan, R., & McCowan, S. (1999). *Embedded Evaluation: Blending Training and Assessment*. Buffalo, New York: Center for Development of Human Services
- Parry, C., & Berdie, J. (1999). Training Evaluation in the Human Services.
- Parry, C., Berdie, J., & Johnson, B. (2004). Strategic planning for child welfare training evaluation in California. In *Proceedings of the Sixth Annual National Human Services Training Evaluation Symposium / 2003* (pp. 19-33). Johnson, B., Flores, V., & Henderson, M. (Eds.). Berkeley, CA: California Social Work Education Center, University of California, Berkeley.
- Pecora, P., Delewski, C., Booth, C., Haapala, D., & Kinney, J. (1985). Home-based family-centered services: the impact of training on worker attitudes. *Child Welfare*, 65(5), 529-541.

# Appendices

## **Appendix A: Child Maltreatment Identification: A Chain of Evidence Example**

*(Cindy Parry & Jane Berdie, Macro Eval Team Meeting September 2004)*

---

Stakeholders rightfully want to know that training “works,” meaning that trainees learn useful knowledge, values, and skills that will translate to effective practice and support better client outcomes. However, many variables other than training can affect whether these desirable goals are achieved (e.g., agency policies, caseload size, availability of needed services). Since it is usually impossible to design training evaluations that control for all of these variables, it makes sense to gather data about training effectiveness at multiple levels. For example, whether the course and trainer meet professional standards, whether trainees found the course valuable, whether trainees learned critical knowledge, whether trainees are able to demonstrate mastery of skills during or after training, and whether agency goals and client outcomes are met (e.g., case plans clearly reflect family involvement in decision making, children receive needed medical care). Data from multiple sources forms a “chain of evidence” about training effectiveness.

This example outlines one of many possible scenarios for what an evaluation of child maltreatment identification might look like at various levels. It is intended to be a concrete illustration of what the steps in the strategic planning grid might look like in one area. The chain of evidence here shows that:

- Trainees were exposed to coursework on the child maltreatment identification competencies.
- The course was content valid and addressed the competencies at the appropriate breadth and depth.
- Trainees found the course relevant and useful.
- Trainees learned specific knowledge needed to identify child maltreatment (e.g., the legal definitions of abuse and neglect, types of burns and breaks associated with maltreatment and those associated with accidents).
- Trainees began to master necessary skills such as being able to:
- Recognize various types of external injuries (cigarette burns, rope marks, splash burns) from slides.
- Recognize common indicator clusters from written case scenarios (e.g., child behaviors associated with sexual abuse).
- Examine a child for injuries.
- Trainees transfer this knowledge and skill to their jobs (e.g., correctly identifying types of injuries, accurately explaining medical reports of injuries, making judgments about child maltreatment based on descriptions of child behaviors and physical evidence).
- Clients achieve better outcomes (e.g. children receive medical care when needed).

Because the intervening variables are not controlled, competing explanations for trainee performance are still possible. For example:

- more workers may be accurately identifying child maltreatment because a particular supervisor emphasizes that rather than because they learned about it in training and/or
- clients may achieve better or worse outcomes for many reasons unrelated to training.

However, evaluation at multiple levels makes a reasonable case for the role of training in an outcome and also allows us to trace back to find where the process may have broken down. For example, trainees may have done well on the knowledge testing but not received enough practice to fully understand and acquire the skill, or trainees may have left training minimally competent in the skill, but were told “we don’t do that here” when they got back to the job.

---

## **Levels of Evaluation**

### ***Tracking***

The task here is to show that new workers have been exposed to the coursework covering the knowledge, skills, and attitudes related to child maltreatment from the CalSWEC competencies and to track that information in a database or series of reports. There are several ways to do this. One might be for each RTA or county providing Core training to develop a master list of what courses in which the CalSWEC objectives are covered. A database could be set up in ACCESS that tracks each new worker’s completion of each course using an identification code that is unique to that worker only, as well as a unique course identification code. The database would also have a table relating the course number to the objective numbers covered in it. Reports could be generated for the State showing numbers of trainees completing coursework in each objective by linking the database tables. These data could be provided in aggregate, but the inclusion of unique IDs for each person allows the county or RTA to know that each trainee is being exposed to each competency and prevent duplicated counts.

### ***Course (formative evaluation)***

The task at this level is to show that the curriculum is content valid and teaches to the competencies and objectives. Curricula could be reviewed to be sure that each competency is being addressed and in sufficient breadth and depth. It is particularly important to determine whether the material related to competencies written at the skill level is really teaching to skill. For example, it would not be enough to teach to the competency, “The worker will be able to identify skin injuries that are usually due to abuse such as, cigarette burns, use of implements (rope, belt, hairbrush) and immersion burns,” by just lecturing while showing slides. There have to be opportunities for participants to practice, e.g., identifying injuries when they see pictures and, conversely, matching descriptions of injuries to pictures.

Other issues which are important at this level (because they set the stage for later levels of evaluation) are:

- the content is spelled out clearly and completely,
- the methods for conducting exercises are written out and
- directions to the trainer are clear.

There are good arguments for the trainer having the flexibility to tailor instruction to the needs of the group. However, this flexibility needs to be carefully balanced against the need for the training to provide a structured and consistent experience that gives all trainees a level playing field against which they will be evaluated.

### ***Satisfaction/Opinion***

At this level, the goal is to show that trainees perceived the course information to be useful. Most agencies offering training already perform this level of evaluation. Its usefulness is extended, however, when items are included related to perceptions of relevance. Items could be included in the end of course feedback forms such as, “I can think of specific cases/clients with whom I can use this training,” or “I will use this training on the job.” Trainees can also be asked to rate the amount they learned in relation to each competency or objective. For example, they can be asked to rate how much they learned about identifying common skin injuries that are due to maltreatment on a scale of 1 to 5 where 1 is nothing and 5 is a great deal. Attitudes may also be assessed at this level.

Usefulness of this information is also extended when ID codes are used to link this feedback with demographics and other types of evaluation. For example, this makes it possible to compare the performance of MSWs to other educational groups. It also helps interpret negative findings at higher levels. For example, if trainees perform poorly on a knowledge test, it may be helpful to know that they didn’t see the information as relevant to their jobs and didn’t expect to use it.

### ***Knowledge***

At this level, the goal is to establish that trainees learned the necessary facts and procedures needed to identify child maltreatment, e.g. “The worker will be able to identify three common behavioral indicators of child sexual abuse in toddlers.” This might be done with a paper and pencil or Classroom Performance System administered test, both pre and post training. Administering the test pre and post-training helps to establish that changes in knowledge occurred as a result of attending the training.

Alternatively, it may be enough to simply know that trainees have the knowledge after training – in this case post evaluation is sufficient once the test has been validated. Key considerations for knowledge tests are content validity (that items cover the important concepts taught) and reliability. A reliable test is necessary to help ensure that what is being measured is actually a change in learning rather than a random fluctuation in test scores. To get adequate reliability, it is usually necessary to administer around 25 items (about a half hour test). A content valid test

includes items that accurately reflect what has been taught, and reflects the relative emphasis or importance of the concepts in the number of questions included on each. To ensure that a test is content valid, it is necessary to have a curriculum that is specific about what information is taught and consistency among trainers in covering the curriculum materials.

## ***Skill***

The goal here is to show that trainees have acquired skill in child maltreatment identification in areas specified by the CalSWEC skill level competencies. At the skill level, evaluation typically is focused on only one or two key skill competencies or objectives because it is too time and resource intensive to try to evaluate all possible skills. In child maltreatment identification, one possibility would be to focus on some key skills that trainees are being taught to do, e.g., recognize injuries that generally do not occur by accident, recognize clusters of behaviors that are associated with maltreatment, and/or be able to examine a child for injuries properly. Trainees could be given case scenario materials as part of an embedded evaluation, and asked to make an assessment of whether child maltreatment has likely occurred and to indicate why.

The assessment would be evaluated using a scoring rubric that specifies how many points to give for each area. For example, a three point scale could be used where no points are given if the area is not addressed at all, 1 point is given if it is addressed but inadequately, and 2 points would be given if it is addressed adequately. Anchors (descriptions and examples of typical 0, 1 and 2 point responses) would be developed and scorers would need to be trained to use the scale reliably.

Typically, this type of evaluation is done as a post test only to save time and reduce the complexity and resource demands that come with trying to develop two versions (for a pre and post evaluation that are different but equal in difficulty). It is also frequently reasonable to assume that child maltreatment identification is not something that trainees come to training knowing how to do well. However, even if done as a posttest only, skill level evaluations typically are complex and require considerable time to develop, test, score, and train people to implement. They may also require considerable curriculum revision.

To do this type of evaluation successfully, some key things need to be in place:

- The skill needs to be taught. Many existing curricula in California and other states are written at knowledge and awareness levels. Several models of teaching to skill exist, but they generally include some version of a lecture to give basic information needed, a demonstration or modeling of what adequate performance of the skill looks like, a chance for trainees to practice the skill and receive feedback under structured conditions, and an opportunity to discuss what using the skill on the job will entail. Some also provide for an additional practice step called independent practice where trainees practice and receive feedback again but with less guidance from the instructors. This is typically the most appropriate place to embed an evaluation component.

- They also require a very structured curriculum or at least a section of curriculum. A “trainer’s guide or outline” doesn’t provide enough structure and consistency on which to base a successful skills evaluation.
- The skill needs to be taught as written in the curriculum by all trainers in order to give each trainee a consistent and fair opportunity to learn the skill. For example, it is not okay for a trainer who doesn’t like to ask people to demonstrate a skill (e.g., the sequence for examining a child) to omit or change the practice step when an evaluation is built from a specific sequence of experiences. These evaluations involve a culture shift for many trainers from being the “sage on the stage” to being more of a guide and coach.
- It takes time and both subject matter and test construction expertise to design and score an embedded skills evaluation task. Scoring is often more subjective and time consuming than with a paper and pencil test. It takes considerable time and expertise to develop a clear, anchored scale. Both time to train people in using the scoring system and time to actually do the scoring are needed.

### ***Transfer***

Evaluations of transfer are most effective if the transfer of learning piece follows from an evaluation done in the classroom. In this way, we can establish that people had a particular skill level when they left training and compare it to the skill level they exhibit on the same evaluation task in the field. In this example, the same scoring rubric developed to score the scenario-based child maltreatment identification in training could be used to score actual child maltreatment identification decisions from the field. If the scoring system is based on more or less universally desirable characteristics, it should be possible to use it successfully even though it is no longer being applied to one case scenario. Consistency of scoring can be an issue, however, particularly if more people are involved and need to be trained to use the rubric (e.g. supervisors). In this case, the consistency issue could be dealt with by having materials de-identified and copied for scoring by a centralized evaluation team.

### ***Outcomes***

This is the most technically demanding level of evaluation and the most costly. In the case of our example, it might be feasible to link to client outcomes by looking at data in case files, or hopefully, in the CWS/CMS information system. For example, it might be possible to look at the achievement of child well-being outcomes related to medical care and reduction of recidivism, and see if the success rates are higher when child maltreatment has been documented in terms of both physical and behavioral indicators.

Evaluation at this level needs some type of comparison group. In this example, since everyone is being trained in child maltreatment identification through Core, it is likely not feasible to compare trained and untrained workers. However, it may be possible to compare case outcomes for those who scored more highly on skill measures at transfer to those who scored lower. Obviously, however, case outcomes would depend on other factors in addition to training. Typically, a large number of cases would need to be included in such a study before a relatively subtle effect could be detected.

**Appendix B: CDSS Common Framework for Assessing Effectiveness of Training: A Strategic Planning Grid**  
**Sample Framework for Common Core**

*(Up to date as of 12/16/04)*

<b>Level of Evaluation</b>	<b>Scope</b>	<b>Description</b>	<b>Decisions Pending</b>	<b>Resources</b>	<b>Timeframes</b>
<b>Level 1: Tracking</b>	Completion of Common Core Training will be tracked	<p>Names of individuals completing Common Core will be tracked by RTAs/IUC and reported to the counties semi-annually. Counties will provide the State with aggregate numbers of new hires and number completing Common Core for the year via the revised Annual County Training Plan.</p> <p>The final Tracking Training report due at the end of FY 2005-2006 will include tracking trainees who complete the Common Core in their respective RTA/IUC/County, which should include, but is not limited to, a minimum of all of the Big 5 content areas.</p> <p>STEC recommended that new hires be required to complete Core training within 12 months.</p>	None	<p>Databases (State, County, RTA)</p> <p>Personnel time to maintain and monitor (both locally and centrally)</p> <p>Protocols for individuals involved in maintaining and submitting data</p>	Tracking to begin by July 2005, in conjunction with roll-out of Common Core Training that includes all of the standardized Big 5 content areas. A final report on Tracking Training data for FY 2005-2006 will be prepared by August 2006.

<b>Level of Evaluation</b>	<b>Scope</b>	<b>Description</b>	<b>Decisions Pending</b>	<b>Resources</b>	<b>Timeframes</b>
<b>Level 2: Course</b>	Big 5	All of the RTAs and counties that provide training will use evaluation data to improve their delivery of Common Core curricula and engage in systematic statewide updates of content and methods. Specific QA procedures will be developed by CDOG.	What type of procedures for observing delivery will be used during Spring 2005 as the Big 5 are piloted?	Personnel time for: Reviewing evaluation results, ensuring quality and consistency e.g. observation/monitoring delivery	BIG 5 curricula will be developed by June 2005. Piloting will begin in March, 2005.  QA procedures for observing delivery will be developed and used during Spring 2005 as the Big 5 are piloted.  Full QA procedures for maintaining and updating curricula will be developed by CDOG during FY 2005-2006.
<b>Level 3: Satisfaction/opinion</b>	Big 5	RTAs and counties that deliver training will continue to collect workshop satisfaction data using their own forms. No standard form will be required due to local constraints related to University requirements of the RTAs. Those that wish to may use a standard form that CalSWEC has developed.  RTAs and counties may use an identification code to link these results with the results of other levels of evaluation, however one is not required.	None	None needed—RTAs/IUC and counties have existing forms and process	Currently being done and will continue.

<b>Level of Evaluation</b>	<b>Scope</b>	<b>Description</b>	<b>Decisions Pending</b>	<b>Resources</b>	<b>Timeframes</b>
<b>Level 4: Knowledge</b>	Big 5	<p>CalSWEC will provide knowledge test items for 5 areas: Human Development, Child Maltreatment Identification, Placement/Permanency, Risk and Safety Assessment and Case Planning and Management, along with supporting research when available. CalSWEC will be responsible for maintaining this item bank and updating or writing new items as curriculum changes. RTAs and Counties are encouraged to submit and review items. Items may be added and validated on an ongoing basis as curricula are updated or new methods of training are implemented.</p> <p>The lead organizations for the Big 5 topics will use the items and research to inform the curriculum development process, and identify areas where new items should be developed. For these 5 areas, essential information presented in training will be standard. The Macro Evaluation Committee has also recommended that a standard core set of items be used in the tests for each area.</p> <p>Macro Eval and CDOG Committees recommend to STEC that as common Big 5 content is developed by each lead agency, each Big 5 lead organization will ID a common item set that must be included in the test during the item validation process.</p> <p>Macro Eval and CDOG Committees recommend that STEC adopt the testing protocol recommended by the Macro Evaluation Committee. Recommendations to STEC include:</p> <ul style="list-style-type: none"> <li>For Big 5 content areas where training occurs over more than 1 day, pre and post tests will be conducted until items are validated. After the items are validated, this decision will be revisited.</li> </ul> <p><i>(continued on next page)</i></p>	<p>What type of data entry (manual, scanned locally, scanned centrally) will be most accurate, efficient and cost effective?</p> <p>Procedures for data transfer from RTAs/IUC or Counties to CalSWEC (depends on method of entry used and will be communicated in final item bank training prior to Spring 2005) during piloting of new Big 5 and related knowledge exams?</p> <p>What should QA process be for piloting process?</p> <p><i>(continued on next page)</i></p>	<p>Staff time to prepare and distribute paper tests.</p> <p>Staff time for data management, input and transfer. CalSWEC is looking into the purchase of scanners to aid in data entry.</p> <p>Cost of scanner or copying/ mailing costs.</p> <p>Staff/consultant time for review and revision of knowledge items Staff/consultant time for analysis and reporting to RTAs, Counties, State.</p> <p><i>(continued on next page)</i></p>	<p>Annotated item bank to RTAs/IUC from CalSWEC by Oct 31, 2004. Lead RTAs/IUC will select items and work with CalSWEC (Leslie, Jane and Cindy) to develop new items or modify existing items and to select final set of items prior to piloting their trainings.</p> <p>Testing will begin in March 2005 with the pilots of the Big 5.</p>

<b>Level of Evaluation</b>	<b>Scope</b>	<b>Description</b>	<b>Decisions Pending</b>	<b>Resources</b>	<b>Timeframes</b>
<b>Level 4: Knowledge (cont'd)</b>	Big 5	<ul style="list-style-type: none"> <li>• For Big 5 content areas where training occurs in just 1 day, only a post test will be conducted until items are validated. After the items are validated, this decision will be revisited.</li> <li>• Recommended time allowed for each test is 45 minutes for pre and 30 minutes for post.</li> <li>• The same test form will be used for pre and post testing.</li> </ul> <p>The RTAs/IUC will conduct evaluation according to agreed upon protocols and send data to CalSWEC to validate items. CalSWEC will validate the items and update the item bank based on the data. After the item validation process, each RTA and/or county will select items from the item bank that fit their curricula and will identify new items to be developed. CalSWEC will work with them to develop these items.</p> <p>Data will be collected from each participant but reported only in aggregate. A confidential ID code that each participant generates from the first 3 letters of mother's maiden name and 2 digits for day of birth will be used to link pre and post tests. Directions will be given for names of less than 3 letters. County ID will be collected but not as part of person ID. A standard set of demographics will be collected and linked with test data (for purposes of item validation/ detecting potential item bias).</p> <p>Participants will be informed of the purpose of the evaluation, confidentiality procedures and how the results will be reported and used. RTAs and Counties will have access to statewide aggregate data and data from their own trainings. RTAs and Counties will use results to determine course improvement and demonstrating knowledge competency of workers in aggregate (not individuals).</p> <p style="text-align: center;"><i>(continued on next page)</i></p>	What should QA process be once the piloting process is complete?	Need to develop tools for QA process during and post piloting process (including tools for observers to use for process and feedback)	

Level of Evaluation	Scope	Description	Decisions Pending	Resources	Timeframes
<b>Level 4: Knowledge (cont'd)</b>	Big 5	<p>A timeline for phases of the evaluation process will be developed and shared with counties to clarify what results will be available to them and how they may be used appropriately.</p> <p>Participants will turn in their tests before leaving the classroom to avoid a loss of item validity if they are circulated.</p> <p>Trainers will have written instructions and/or training in how to administer and debrief evaluations and monitor the ID process.</p>			

Level of Evaluation	Scope	Description	Decisions Pending	Resources	Timeframes
<b>Level 5: Skill</b>	Child Maltreatment	<p>For one identified trial skill, Child Maltreatment Identification, the content and method of delivery will be standard. All RTAs and counties that deliver training will integrate the curriculum and embedded evaluation under development for this skill into existing Core.</p> <p>Feedback will be used for course improvement and demonstrating competency of workers in aggregate (not individuals). Data will be collected from each participant but reported only in aggregate. Responses will be confidential. The personal ID and information from the Demographics form (discussed in Knowledge) will also be used for skill level evaluation. Participants will be informed of the purpose of the evaluation, confidentiality procedures and how the results will be reported and used.</p> <p>Evaluation will use case scenarios and slides. It will focus on identification of whether physical abuse as defined by CA W&amp;I Code has occurred. Evaluation will be post training only, because of the time involved in evaluating skill/application of knowledge.</p> <p>Participants will turn in any evaluation forms before the trainer processes the evaluation exercise and will not take any evaluation materials out of the classroom.</p> <p>Trainers will have written instructions and/or training in how to administer and debrief evaluations.</p> <p>RTAs and Counties will have access to statewide aggregate data and data from their own trainings. RTAs and Counties will use results to determine the extent to which skill acquisition is occurring.</p>	<p><b>Southern RTA (lead org for CMI) will make a recommendation concurrent with curriculum development re: whether or not to include neglect in the skill evaluation.</b></p> <p><b>Final content of evaluation tasks and administration and scoring procedures?</b></p> <p><b>Procedures for data transfer from RTAs or Counties to CalSWEC?</b></p> <p><b>Development of a minimum competency standard against which to judge performance?</b></p>	<p>Trainer/SME time for consulting on design and scoring rubrics</p> <p>Trainer time for learning administration and debriefing of evaluation</p> <p>Staff/consultant time for evaluation design, analysis and reporting to RTAs, Counties, State</p>	<p>Southern will work with CalSWEC (Jane and Cindy) to develop the knowledge/application of skill test to be ready for piloting along with the curriculum in March, 2005</p>

<b>Level of Evaluation</b>	<b>Scope</b>	<b>Description</b>	<b>Decisions Pending</b>	<b>Resources</b>	<b>Timeframes</b>
<b>Level 6: Transfer</b>	TBD	Suggested transfer of learning activities will be developed for the Common Core as part of the Big 5 curriculum development process. Evaluation of these activities (and/or transfer generally) is for future development, once the mentoring study is completed and recommendations are reviewed.			Phase 1 of the mentoring study projected to be completed by April 2005. Preliminary feedback to inform supervisor curriculum development will be available by the end of October.
<b>Level 7: Agency/client outcomes</b>	TBD	Using the entire framework, California can begin to build the 'Chain of Evidence' necessary to evaluate the impact of training on outcomes.			Evaluation on specific outcomes can be developed as the state and federal outcomes are refined and the framework is implemented and refined.

## Appendix C: Big 5 Curriculum Development and Evaluation Considerations

---

### **Sections:**

**Part A: Types of Evaluation & Possible Training-Related Activities**

**Part B: Time Frames**

**Part C: Example of Curriculum after Reviewing Items**

---

### ***PART A: Types of Evaluation and Possible Training-related Activities***

#### **1. Satisfaction/opinion (happiness sheets)**

→ Identify key competencies so that the evaluation can address participant opinion about the quality of training or how much was learned related to these competencies.

#### **2. Knowledge**

→ Within the various competencies and learning objectives, identify the **detailed** key knowledge that is covered in the training so that relevant items can be selected from the existing Item Bank or, if these items do not exist, they can be newly developed.

→ Review Item Bank items to see if there is content that training could/should be covered that is currently not covered. (See Part C for an example).

→ Review CalSWEC literature reviews for content that might be included in curriculum. *A compilation of the literature reviews was distributed at the September 2<sup>nd</sup>, 2004 Content Development Oversight Group meeting.*

→ Work with Cindy and Jane to identify new areas where knowledge items should be developed or current items can be improved.

#### **3. Embedded Evaluation of Skill (recognition of skill or demonstration of skill)**

→ Determine what specific skill will be both taught and evaluated (limit to one or two skills because embedded evaluation takes time and resources within the training day). The skill should be one that is both critical for effective practice on the job and one which can be measured in a classroom setting.

→ Determine what level of skill will be evaluated. Generally, there are two levels: 1) ability to recognize when a skill is being done correctly by another person or 2) ability to conduct the skill oneself. Clearly, the second is a more useful measure of skill acquisition but it requires a higher level of resources (time, evaluation tools, and often on-site evaluators) than the first.

- Ensure that the curriculum is written at the right level to support learning the skill at the level at which it will be evaluated (can recognize the skill being done correctly or can perform the skill) - the content is both focused and in-depth and the training methods enable the participants to learn the skill, covering the five steps of skill curriculum:
  - ***Explain:*** *is the relationship between the skill and knowledge made sufficiently clear?*
  - ***Model/demonstrate:*** *is the skill demonstrated (e.g., by the trainers or on videotape) so that participants see the skill done correctly?*
  - ***Practice:*** *are there one or more structured, standardized practice sessions?*
  - ***Feedback:*** *is there a structured approach for participants to receive feedback, preferably from the trainer and/or other skilled person? If participants give feedback, have they been trained to do so?*
  - ***Discuss transfer:*** *is there a structured approach for discussing use of the skill on the job?*
  
- Ensure that adequate time is available in the training day to teach the skill.
  
- Ensure that trainers are prepared to teach the curriculum at the skill level. For curriculum writers, this means including specific instructions for the trainer and preparing support materials, e.g., case studies, dialogues, and videotapes to demonstrate a skill, structured case materials to practice a skill and to guide feedback, and trainer notes on common issues that come up during practice and feedback sessions. It might also include guidelines for trainer support, e.g., how to help trainers become proficient in training this curriculum at the skill level.
  
- Ensure that the curriculum writers and evaluation designers are working together as soon as the skill is selected. The evaluator will work with the curriculum writer to design the protocol (and all of the tools needed to conduct the evaluation) and will then analyze the evaluation data. Some of the considerations will be:
  - *What will the evaluation exercise be (extension of training activity or new activity; using slides, scenarios or a combination; how much time; how many aspects/items will it include)?*
  - *What procedures will be used (who will administer the evaluation, who will score the participants' performance on the exercise)?*
  - *What will adequate performance look like?*
  - *What kind of instrument will best capture the participant's performance (i.e., the rating sheet)?*
  - *How will performance be rated (e.g., what will be acceptable, outstanding, or inadequate)?*
  - *When and how will feedback be provided to participants?*
  - *What instruments and instructions are needed for the participants?*
  - *How will data be collected and analyzed?*

- Ensure that the right resources are available to conduct the evaluation. This might mean experts in the training room to evaluate performance and give feedback (e.g., if participants are practicing interview skills) or it might mean an expert who could evaluate performance later (e.g., from written documents if participants have completed a case plan or a risk assessment or from video if participants' practice sessions were videotaped).

#### **4. Transfer of Learning**

- Curriculum directs the trainer to explain what skill will be assessed on the job and how it will be assessed.
- Training materials include job aides to support participants using the skills.
- As above, the curriculum writer and evaluator will need to work together.

---

### ***PART B: Time Frames***

- If you plan to use evaluation results to critique and modify curriculum, consider the first few deliveries as pilots. It will take some time to get enough evaluation results back to know if changes in the curriculum are warranted.
- If you are working with an evaluator, e.g., for a skill level curriculum, you will need to build in time to the development of the training to work together on the evaluation.
- If you are developing a skills module, keep in mind that the modeling/demonstrating, the structured practice, and the structured feedback phases will take much more time during the training day than the more typical type of training in which there is no modeling and only simplified practice and minimal feedback. Also, unless you can fully rely on experts (such as seasoned workers or supervisors) to be at the training and to observe the participants' practice and give feedback to them, you may need to build in time to teach participants how to observe and give feedback (generally the skills of new staff such as those in Core training are not adequate for this).

### **PART C: An Example of Revising Curriculum after Reviewing Items**

**Item #CM022 is as follows: Which description of an injured child is most likely to lead to suspicion of child abuse?**

- a. LaTasha's mother rushes into the emergency room holding her toddler. LaTasha has second degree burns on her shoulder and upper arm. Her mother explains that LaTasha was playing in the kitchen and stumbled into the ironing board. The iron fell and struck LaTasha on the shoulder and slid off her upper arm.
- b. Jorge, age 6, is brought to the medical clinic for a persistent cough. The health practitioner observes circular, reddened burn marks on the child's back. The grandmother, a Mexican immigrant, refuses to discuss the red marks but says through an interpreter "she's done everything she can do to make him better."
- c. **Benjamin's mother picks up her son, age 15 months, from his baby-sitter. Benjamin is crying and his feet and ankles are red and blistered. The baby-sitter says that Benjamin stepped into hot water she was using to wash the kitchen floor. She says she has been putting ointment on him.**
- d. Mark's parents call 911 because Mark, age 2, was burned by an old radiator in the house. Mark had tried to reach for a toy under the radiator and had inserted his arm through a narrow slot in the heater. Mark has burns on both sides of his arm.

\*\*\*\*\*

This item taps the “understanding” level of learning in that it requires the respondent to be knowledgeable about several issues and how they interface, e.g., types of injuries and probable causes, age stage behaviors of children, typical triggers for adult anger, cultural healing practices, and common sense knowledge about “how things work” (such as how easily an iron can be knocked off an iron board).

The content that would likely be covered in training for this item would include both information about each of these issues and the interface. The information about the issues would include:

- Types/sites of injuries that are usually accidental except on children who are not yet mobile (such as abrasions on the knees and elbows)
- Types/sites of injuries that may be accidental or non-accidental but which are almost always non-accidental in children who are not yet mobile (spiral fractures of limbs)
- Types of injuries that are almost always non-accidental (immersion burns, cigarette burns)

- Typical and atypical child behaviors by age (inquisitive 2 year olds explore with very little sense of safety, children of any age whose skin is beginning to burn pull away if they are able - they do not continue to touch the source of heat)
- There are normal, healthy childhood behaviors associated with development that are triggers for parental abuse and there are also common abusive patterns of parental response to these behaviors, e.g.:
  - colic babies cry and parents shake them,
  - toilet training toddlers soil more than they succeed, parents want to clean them, and an angry parent may immerse the child in hot water,
  - latency age children often day dream and dawdle and frustrated parents may hit them
- There are cultural behaviors for healing that leave marks on children, e.g.,
  - “cupping” (hot cups applied to back or chest when child has respiratory ailment)
  - subdural hematoma of the fontanelle caused by sucking (baby’s soft spot sinks due to dehydration when ill and parent sucks it to bring it up – baby actually just needs to be hydrated)
- Children whose skin is darker than most Caucasians may have birth marks that appear to be bruises to Caucasians who are not familiar with this phenomenon.

To address the interface, the curriculum could present a number of scenarios (much like those in this item) and opportunities for the trainees to analyze them. The use of slides to show injuries would be helpful. Quotes from parents that demonstrate how and why they use cultural healing methods could be helpful.

## **Appendix D: Teaching for Knowledge & Understanding as a Foundation for Evaluation**

*(Cindy Parry & Jane Berdie, Macro Eval Team Meeting April 2004)*

---

### **I. Considerations for Developing Curriculum for Knowledge Competencies**

#### **A. Statewide Decisions: Deciding what to include (Content that is...)**

- Relevant to practice needs
- Accurate and up to date
- Consistent with law and policy
- Balanced (both sides of controversies need to be presented and opinion distinguished from factual or evidenced based information)
- At the right level to develop the competency or meet the stated objective (e.g. with enough breadth and depth of information)
- At the right level to provide pre-requisite or foundation knowledge for later skill or knowledge development

#### **B. Regional/County Decisions: Structuring the presentation...**

- Limit lecture, PowerPoint or other presentation of information to no more than 20 minutes
- Provide opportunities to check for and deepen participant understanding
  - Discussion
  - Examples and stories
  - Quizzes
  - Opportunities to apply new information
- Child Welfare information can have powerful affective component, plan to deal with emotional content if necessary

### **II. Additional Considerations for Developing Curriculum When Knowledge will be Evaluated**

- The key to successful evaluation is a match between what is taught and what is tested.

#### **Regional/County Decisions:**

<b>Curriculum:</b>	<b>Selecting Test Items:</b>
<ul style="list-style-type: none"> <li>• Must spell out all key points and relevant information to be taught</li> <li>• Must specify timeframes &amp; method(s)</li> <li>• Delivery must be consistent and follow the curriculum</li> </ul>	<ul style="list-style-type: none"> <li>• Must reflect key points</li> <li>• Must reflect range and relative importance of information taught</li> </ul>

## **Appendix E: Writing Curricula at the Skill Level as a Foundation for Embedded Evaluation**

*(Cindy Parry & Jane Berdie, Macro Eval Team Meeting April 2004)*

---

### **PART A: A MODEL FOR TRAINING TO THE SKILL LEVEL**

Teaching skills in child welfare work involves the integration of competencies at various levels, including:

- *Knowledge* (e.g., about child abuse dynamics and laws and agency procedures)
- *Cognitive strategies on how to apply knowledge*, i.e., using knowledge to guide behaviors/actions. An example is considering/weighing information in light of a theory of behavior or a framework of practice.
- *Behaviors or actions*. Typical behaviors/actions in child welfare include assessing, planning, documenting and decision-making. Many behaviors/actions are inter-personal (such as interviewing, participating in planning meetings, testifying in court). Others are individual, e.g., observing, reading records, and documenting.

Different skills involve various mixes of all of these competencies. Whatever the mix, a useful way to train to skills is with the following steps:

<b>Step 1</b>	<b>Explain (and discuss)</b>
<b>Step 2</b>	<b>Demonstrate (and discuss)</b>
<b>Step 3</b>	<b>Practice</b>
<b>Step 4</b>	<b>Feedback</b>
<b>Step 5</b>	<b>Discussion</b>

These steps need not always be completely sequential. For instance, you might want to go back and forth between explaining and demonstrating if there are several aspects to the skill.

---

### **PART B: SKILLS TRAINING 'STEPS'**

#### ***STEP 1: EXPLAIN (& DISCUSS)***

In writing curricula, keep in mind what the behavior/action should look like and what knowledge and cognitive strategies go into it (these we often frame as learning objectives).

*Formative criteria include the following:*

- Are the competencies and objectives at the right level and are they both comprehensive and specific enough to “nest” the information we want to impart?
- Is the information accurate, as well as sufficiently comprehensive and specific to the roles of the trainees?
- Is the information imparted during training in ways that support interest and learning by a range of learning styles (including time for discussion and clarification)?
- Is the relevance of the information to the focal skill made clear?
- Are the key points emphasized so that the trainees are ready to focus on them during step two when the skill is demonstrated?
- If there is an assumption that trainees already have some or all of the information (e.g., from previous training), is that confirmed/reinforced? Is the information reviewed at the right level of detail during this step?

## **STEP 2: DEMONSTRATE (& DISCUSS)**

Demonstration of a skill helps people to see what the skill actually looks like. Often this component of skill training is minimized or even left out. Sometimes it is covered but only in the negative e.g., “watch this interview (on tape or by trainers) and critique what is wrong with it.”

*Demonstration can be provided in a variety of ways including:*

- Video or audiotapes (audio tapes work well for training phone screeners). There is a dearth of good tapes and so we often have to use them in a negative/positive mode (“some of what you will see/hear is good practice, while some could be better”).
- Trainers demonstrate the skills. A demonstration with two trainers is better than one with a trainer and trainee, but often we try to involve a trainee in the role play. It is much more effective for two or more trainers/coaches to demonstrate using a script – this insures that the skill will be demonstrated as you want it to be.
- Written material – a case plan and/or a case study.

*Formative criteria include:*

- Is the demonstration primarily positive, i.e., shows the skill(s) as we want them to be done?
- Does the demonstration cover and highlight the key components of the skill?
- Does the demonstration attempt to exclude extraneous information so that the trainees can focus on the important components?
- Does the content, timing and method of discussion help trainees to identify the following:
  - Key relationships between knowledge, cognitive strategies, and action/behaviors
  - What makes for an effective use of skill
  - Common barriers to effectiveness and strategies to overcome

- What might be variations of trainee approach and style that are congruent with the skill

### **STEP 3: PRACTICE**

This step is the opportunity for trainees to practice the skill. Clearly, in the classroom practice is in a hypothetical situation and usually only one portion of the skill can be practiced, and only for a limited time.

This step needs to be highly structured so that the key components can be practiced and so that the next step (feedback) can be useful. It is necessary that trainees have an opportunity to practice the skill before the embedded evaluation.

*Formative criteria include:*

- ➔ Is the practice focused on the key components of the skill?
- ➔ Are there sufficient directions and support materials to target and standardize the practice?
- ➔ Is there a component built in to the practice that creates a role for someone (preferably a trainer or coach but at minimum another trainee) to observe and give feedback and is there guidance on this role so that the feedback is meaningful?
- ➔ Is there sufficient time given to allow each trainee practice?

### **STEP 4: FEEDBACK**

This is the step that gives each trainee the opinion of the person(s) charged with the role of observer/critiquer.

*Feedback is most effective when it:*

- ➔ Targets the key skills
- ➔ Is individualized, i.e., each trainee receives feedback about his/her practice
- ➔ Identifies both strengths and challenges of each trainee, giving examples illustrating actual behaviors
- ➔ Helps each trainee integrate knowledge, cognitive strategies, and behaviors
- ➔ Gives suggestions for improvement and strategies that can be implemented on the job
- ➔ Is both oral and in writing
- ➔ Uses a standardized written format that identifies items, strengths, challenges, suggestions for improvement, and strategies for transfer.

*Formative criteria include:*

- ➔ Is there a format for giving feedback that includes a standardized approach and instrument?

- Is there an opportunity for trainees to receive feedback from a trainer or coach? Have the trainers/coaches been taught to perform this skill?
- If the feedback will be provided partly or solely by other trainees, have trainees been prepared to give feedback? This can be done by training them to do so early in the training, i.e., by including a mini-training on observing and giving feedback. This mini training would itself be a skills training: it would include the steps of Explanation, Demonstration, Practice, Feedback and Discussion of Transfer (“transfer” being conducting feedback as part of training exercises). Thus, the trainer would:
  - Explain and then demonstrate observation and feedback
  - Provide an opportunity for trainees to practice giving feedback to each other
  - Facilitate discussion of their experience and how they will use this to conduct feedback throughout the rest of the training (combining the steps of Feedback and Discussion). This would provide some preparation for trainees in giving feedback to each other during the subsequent skills building sessions.
- Is sufficient time given for feedback?

### **STEP 5: DISCUSSION**

This last step is an opportunity for the group as a whole (or small groups) to discuss the skill, the practice session, and what might be transfer implications. Again, the discussion session should be sufficiently structured to cover these issues.

*Formative criteria include:*

- Do the instructions to the trainer provide sufficient detail to cover the key issues:
  - Review of the skill
  - Experiences in practice including typical strengths, challenges, obstacles and strategies
  - Transfer implications, e.g., barriers and supports to using the skill on the job

**Appendix F: Bay Area Trainee Satisfaction Form**  
**(Next Page)**

# BAY AREA ACADEMY

## Course Evaluation

**Workshop Title:** \_\_\_\_\_

**Date:** \_\_\_\_\_ **Trainer:** \_\_\_\_\_

**Invited Counties:** \_\_\_\_\_

**Your Unique ID:** \_\_\_\_\_

For each question, please check the box under the number that best represents your assessment of the course. Your assessment of this training event will help us plan future Bay Area Academy training programs. Thank you!

**EFFECTIVENESS OF COURSE LEARNING OBJECTIVES** (see course flyer)

	Strongly disagree				Strongly agree	
	1	2	3	4	5	Didn't cover
1. The Learning Objectives were made clear to me	<input type="checkbox"/>					
2. The Course was consistent with the stated Learning Objectives	<input type="checkbox"/>					
3. All of the Learning Objectives were met	<input type="checkbox"/>					
4. The Course covered issues of permanency	<input type="checkbox"/>					
5. The Course covered issues of safety	<input type="checkbox"/>					
6. The Course covered issues of child well being	<input type="checkbox"/>					

**CULTURAL APPROPRIATENESS OF COURSE MATERIAL** (ethnicity, race, class, family culture, lifestyles, language, sexual orientation, physical and mental abilities)

	Strongly disagree				Strongly agree	
	1	2	3	4	5	Didn't cover
7. Exercises/Discussion/Handouts included material on cultural diversity	<input type="checkbox"/>					
8. At least one Learning Objective in the course material applied to more than one cultural group	<input type="checkbox"/>					
9. Overall, Course material appropriately addressed examples of issues of cultural differences	<input type="checkbox"/>					

**EFFECTIVENESS OF COURSE TRAINER**

	Strongly disagree				Strongly agree	
	1	2	3	4	5	Didn't cover
10. Provided a well-organized presentation	<input type="checkbox"/>					
11. Communicated material in clear and simple language	<input type="checkbox"/>					
12. Provided appropriate examples relevant to Child Welfare	<input type="checkbox"/>					
13. Trainer motivated me to incorporate new ideas into practice	<input type="checkbox"/>					
14. I would recommend this training to a co-worker	<input type="checkbox"/>					

**EFFECTIVENESS OF PRESENTATION**

	Not effective				Very effective	
	1	2	3	4	5	Didn't use
15. Material was presented in multiple formats:						
a) Lecture	<input type="checkbox"/>					
b) Facilitated discussion	<input type="checkbox"/>					
c) Small group breakouts	<input type="checkbox"/>					

- |   |                          |                          |                          |                          |                          |                          |
|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| d) Role Plays                           | <input type="checkbox"/> |
| e) Case Examples                        | <input type="checkbox"/> |
| f) Technology - video, PowerPoint, etc. | <input type="checkbox"/> |

Not  
effective

Very  
effective

**OVERALL RATINGS**

Please rate the trainer and course on a scale of 1 to 5, where 5 is the highest rating

**1      2      3      4      5**

- |  |                          |                          |                          |                          |                          |
|--|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| <b>16. Overall Rating of the Trainer</b> | <input type="checkbox"/> |
| <b>17. Overall Rating of the Course</b>  | <input type="checkbox"/> |

**Participant Comments:**

18. What aspects of today’s training were most helpful for you? Why?

19. What aspects of today’s training need to be changed or improved? Why?

20. How will you apply what you’ve learned in this workshop to your job? Please provide at least two specific examples.

21. Was this training a good use of your time? Please explain.

22. Additional comments:

## **Appendix G: Protocol for Building and Administering RTA/County Knowledge Tests Using Examiner and the Macro Training Evaluation Database**

*(Cindy Parry & Jane Berdie, Macro Eval Team Meeting September 2004)*

---

This is a preliminary set of steps for talking purposes. In some steps, for example, submitting the test data, the exact procedures will depend on whether or not data is scanned locally or manually entered. A complete protocol will be provided with the item bank database.

*Note: CDOG recommended that Big 5 content area lead organizations choose a standard set of items for knowledge tests in each of their content areas for the piloting phase of the process.*

---

**Step 1**      **Identify content you want to test at the knowledge and understanding levels of learning within a Big 5 area.**

**Step 2**      **Select Items from the Macro Training Evaluation Database.**

- The database in Examiner format will be provided by Leslie and Cindy. Select from existing items. **Do not modify any of the items in the Macro Training Evaluation Database.** We will be collecting additional feedback for modifications after the first round of item analysis. We will make all changes at one time and forward the revised item bank database to you.
- **Do not change the CalSWEC number of each item.**
- If you want to add locally developed items to your knowledge test in addition to the item bank items, we will provide a series of item numbers for each RTA/county to use to designate their own items. You may e-mail these items to Leslie if you want them to go through the review process for inclusion in the item bank, or keep them as local items only. These items will not be included in the analysis unless you request that they be reviewed for inclusion in the item bank.
- We have not set a formal minimum number of items to be selected for each topic area test but keep in mind that for reliability and validity a 30 item test is best. Tests with fewer than 20 items are not recommended.
- **Do not randomize or otherwise change the answer options labeled a, b, c, or d, on multiple choice items at this time.** You may randomize the order in which the items (i.e. the questions) are presented.
- If you are using a pre and posttest you have two options: 1) use the same items pre and post. However, we encourage you to, at minimum, change the order of the items from pre to post. 2) Use different but overlapping sets of items for the pre and the post. If you use different sets, overlap them by one-half to one-third (e.g. for a 30 item test, make 10 to 15 items the same on the pre and the post). The overlap is needed for the pilot phase of the project,

and will no longer be necessary when the items are fully validated and scaled.

**Step 3**      ***Print your customized knowledge test.***

At the March 28, 2003, Macro Evaluation Team meeting, the group agreed that test packets would consist of:

- *A Cover page*
- *The informed consent page*
- *The ID Code Assignment and Demographic Survey*
  - *A Participant ID:* An ID code is necessary to store, link and track data within Examiner, e.g. to combine demographics with test responses in order to look for item bias. It also allows us to answer questions like “Do MSWs learn in Core training or do they already know the information?” and it is necessary to link pre and posttests. The proposed ID system is to use a code that the trainee would generate based on immutable personal data that others would not know. CalSWEC has used a system where the ID is the first 3 letters of mother’s maiden name and a 2 digit day of birth.
  - *A site ID* (probably using 01-58 for the counties and sub codes for the academies)
  - *Demographics* using the CalSWEC demographic form
- *The Directions to the Training Liaison*
- *The test items*

Examiner generates a test number. This test number tells Examiner what key to use to score the test. This number should appear on all printed tests. If you are using bubble sheets, the test number must be “bubbled” in on the answer sheets. Paper tests and answer sheets should also carry the date and if applicable a code for pre or post test.

**Step 4**      ***Administer the Test.***

- A list of talking points to be used to explain the test to trainee will be provided to each RTA/county administering the knowledge tests.
- Allow about a half hour to introduce and administer a 25-30 item test (it takes about a minute to read and answer a multiple choice item - longer for scenarios)
- Trainers, RTCs or others responsible for administering the tests should check for ID codes on all test forms and be sure to collect all test forms before participants leave the classroom. If the items leave the classroom and circulate, the validity of the tests will be compromised.

**Step 5*****Submitting the test data***

- *Collecting and transferring test data to CalSWEC*
  - Scanning versus entry of paper forms
  - Developing customized report options with Examiner
  - Resource needs

**Step 6*****Getting results back***

- *The items*
  - As the pilot proceeds, items will be tested. Individual items will be dropped, modified or added on a set schedule and a new copy of the item bank will be sent to you (schedule to be discussed). If an individual item is clearly not working as intended between update cycles, you will be notified via e-mail and asked to mark it as “unselectable” in Examiner.
- *Your RTA or county*
  - Schedule to be discussed
- *Statewide aggregate*
  - Schedule to be discussed
  - During the pilot test results may not reflect the performance of the group as much as the performance of the test items, thus county, regional, and statewide reports may not be immediately available.

**Appendix H: Standardized ID Code Assignment & Demographic Survey  
(Next Page)**

One Digit RTA Code: \_\_\_\_  
Two Digit County Code: \_\_\_\_  
Identification Code: \_\_\_\_  
(see below)

**STANDARDIZED CORE CURRICULUM**  
**Identification Code Assignment & Demographic Survey**

---

**Dear Standardized Core Curriculum training participant,**

Be sure to fill out the following survey once during the course of the SCC training. If you have already filled out this form, you do not need to do so again.

**BEFORE YOU BEGIN...**

**YOUR IDENTIFICATION CODE:**

In order for us to track your evaluation responses while maintaining your anonymity, we need to assign you an *identification code*. We would like you to create your own *identification code* by answering the following questions:

1. What are the first three letters of your mother's *maiden* name? (Example: If your mother's maiden name was Alice Smith, the first three letters would be: **S M I**. **If the name has less than three letters, fill in the letters from the left and add 0 (zero) in the remaining space(s) on the right.**

\_\_\_\_\_

2. What are the numerals for the DAY you were born? (Example: if you were born on November 29, 1970, the numerals would be **29**). If your birth date is the 1<sup>st</sup> through the 9<sup>th</sup>, please put 0 in front of the numeral (example: **09**).

\_\_\_\_\_

Combine these numbers to create your identification number (example: **SMI29**). Please write your identification code in the space at the top right corner of this questionnaire. *Remember your identification code and write it at the top of every evaluation form provided to you throughout this training.*

---

**DEMOGRAPHIC SURVEY:**

By providing us with the following demographic information you will be helping us to understand the effectiveness of this training for future participants. Your participation with this survey is completely voluntary and all of the information will be kept entirely confidential. The information you provide us will not be associated with your identity or your performance in any way.

**1. What is the highest level of your formal education? (Check appropriate space(s) below.)**

- |                                       |                                       |  |
|---------------------------------------|---------------------------------------|--|
| <input type="checkbox"/> High School  | <input type="checkbox"/> BSW degree   | <input type="checkbox"/> PsyD                                |
| <input type="checkbox"/> Some college | <input type="checkbox"/> MA/MS degree | <input type="checkbox"/> PhD – Field related to social work? |
| <input type="checkbox"/> BA/BS degree | <input type="checkbox"/> MSW          | <input type="checkbox"/> Yes <input type="checkbox"/> No     |

One Digit RTA Code: \_\_\_\_  
Two Digit County Code: \_\_\_\_  
Identification Code: \_\_\_\_

**2. How long have you been working in the field of child welfare?** \_\_\_\_ years \_\_\_\_ months  
(a public or private agency whose client population is primarily part of the CWS system)

**3. How long have you been in your current position?** \_\_\_\_ years \_\_\_\_ months

**4. Did you participate in the Title IV-E program, which offers stipends to BSW/MSW candidates who specialize in public child welfare, or in a state or county stipend program?**

YES (please answer questions below)       NO (skip to question 9 below)

*IF YES....*

→ *In which program did you participate?*

IV-E (LA DCFS)    ?     IV-E (CalSWEC)     Other state

→ *Were you in the child welfare field prior to your Title IV-E participation?*

YES (please answer question below)     NO (skip to question 9 below)

*IF YES....*

→ *What kind of child welfare position did you have prior to your Title IV-E participation?*

VOLUNTEER       PAID

**5. Do you hold a current license as a mental health practitioner?**       Yes       No  
If yes, which one?     LCSW     MFT     Other \_\_\_\_\_

**6. How do you identify yourself in terms of ethnicity/race? (Check the appropriate space below)**

African American                       Hispanic/Latino  
 American Indian/Alaska Native     Multi-racial (specify): \_\_\_\_\_  
 Asian/Pacific Islander                 Other (specify): \_\_\_\_\_  
 White/Caucasian

**7. What is your age?** \_\_\_\_\_ years

**8. What is your gender?**     Female     Male     Other (specify) \_\_\_\_\_

**9. In what language are you most comfortable reading and writing?** \_\_\_\_\_

**10. Do you currently carry a caseload?**       Yes     No

**11. If yes, how many cases?** \_\_\_\_\_

**Thank you for your assistance.**

# **Appendix I: Moving from Teaching to Evaluation—Using Embedded Evaluation to Promote Learning and Provide Feedback**

*(Cindy Parry & Jane Berdie, Macro Eval Team Meeting April 2004)*

---

## **EMBEDDED EVALUATION FAQs**

### ***1. What is Embedded Evaluation?***

Embedded evaluation uses exercises that are built into the training day, both to promote learning and to provide evaluative feedback. Designs for embedded evaluations range on a continuum from the relatively simple to the more complex. For example, the trainer might observe performance of a role-play and record performance on one or two key objectives on a checklist, or it might be more complex, where an outside evaluator administers the interview task with a set script and a trained actor playing the part of the client. Less complex embedded evaluation is most useful for providing feedback for course improvement and giving trainers an idea of how the class in general is picking up on key points. More complex embedded evaluation is required when the goal is to document (or certify) that individuals have met a specified standard of competency.

### ***2. What is Embedded Evaluation used for?***

Embedded evaluation is most often used to evaluate skill-based competencies. Skill-based competencies are competencies that define a desired behavior, activity or interaction; for example, interviewing a child, assessing risk, identifying indicators of child maltreatment, writing a court report, writing a case plan, etc.

Embedded skills evaluation often involves the observation of a behavior in the training room. However, it could also involve the evaluation of written materials if the skill being taught is to produce a written product such as a court report or case plan. It might also involve making judgments based on slides and written scenario materials when demonstrating a skill like assessment. For obvious ethical and practical reasons, real children and families can't be present in the classroom. However, reasonable substitutes for skill demonstration are available, and include assessing risk from a written scenario, simulated initial reports, interview transcripts, safety assessment forms, or using slides to identify injuries possibly due to physical abuse. What is important is that the evaluation task mirrors the on-the-job use of the skill as closely as possible.

### ***3. What is an example of Embedded Evaluation?***

Embedded performance tasks may be thought of as exercises that are a part of the training as well as being an evaluation method. For example, trainees might be assigned a case planning exercise based on a set of written scenario materials as part of their instruction. The evaluator

would work with the curriculum developers and trainers to identify the key points that should be addressed in the plan and develop a scoring rubric that would be used to assess how well each trainee met the objectives for that exercise. Trainees' scores would then be analyzed and reported back to the evaluation's stakeholders, such as the training program administrator(s), curriculum developers, trainers, trainees and their supervisors, or others.

#### **4. What are the roles of the Curriculum Developer, Training Administrator, Trainer and Evaluator?**

##### **A. The role of the trainer, curriculum developer, or subject matter expert is to:**

- ⇒ Advise on the design of the task and administration logistics.
- ⇒ Help identify the dimensions of competent performance (items).
- ⇒ Identify what competent performance on each dimension would look like (anchors).
- ⇒ Make recommendations about level of overall performance needed for competency (how many items should someone be expected to answer correctly).
- ⇒ Help conduct the evaluation. The trainer is usually the person who sets up and runs the evaluation exercise in the classroom. Training administrators and others may provide assistance with logistics (e.g. arranging for a second trainer if needed to run the exercise, providing for assistance with classroom technology)
- ⇒ Help score the trainees' responses. The trainer would need to participate in scoring if his or her expertise is needed to judge the adequacy of an open-ended or behavioral response.

##### **B. The role of the evaluator is to:**

- ⇒ Structure the task with the trainer and curriculum developer so that the desired feedback can be obtained,
- ⇒ Develop the evaluation design and evaluation instruments,
- ⇒ Conduct the evaluation (usually with the help of the trainer(s),
- ⇒ Analyze the evaluation data,
- ⇒ Write evaluation reports, and
- ⇒ Consult with and provide information to the curriculum developers, trainers, training administrators and other subject matter experts.

Collaboration between the training developer and evaluator is critical to the success of embedded evaluation. The training developer and evaluator jointly develop and agree upon the design of embedded evaluations.

#### **5. What can Embedded Evaluation provide?**

Embedded evaluation either builds on existing exercises or designs new tasks that can be used as both instructional and evaluation opportunities. This linkage enhances trainee learning and provides feedback to trainers for course improvement, while also providing important data on

trainees' acquisition of skills. Embedded skill evaluations in the classroom are promising for two additional reasons.

First, skill level evaluation tasks are time consuming and logistically difficult. Within the training day, an evaluation task that wasn't integrated with instruction would take too much time away from an already tight schedule. Building on existing exercises or designing new tasks that can be used as both instructional and evaluation opportunities is efficient, and provides the added value of integrating and enhancing both trainee learning and the evaluation data.

Second, using embedded evaluations during training provides a baseline for linking training performance with transfer activities. One necessary prerequisite to transfer of learning to the job is initially having learned the skill. Embedded evaluation can help to document to what extent that learning is taking place in the classroom and to what extent transfer could reasonably be expected to take place even under optimal conditions in the field.

### ***6. What are the steps in designing an Embedded Evaluation?***

Embedded evaluations usually follow the same general sequence of steps in design and implementation:

1. Consult with stakeholders regarding purpose and desired outcomes of the evaluation.
2. Identify an appropriate competency or competencies to be the focus of the evaluation.
3. Review curricula and observe a course session to determine whether the curriculum supports skill development sufficiently or if modifications are needed.
4. With the curriculum developers and trainers, select a skill development exercise to become the basis for the evaluation. Alternatively, an exercise may be developed jointly by the curriculum developer/trainer and evaluator to address a particular competency.
5. Design the evaluation making modifications to the exercise selected if needed.
6. Design the scoring rubrics, assessment instruments or surveys and procedures to be used to collect evaluation data.
7. Pilot test the evaluation for reliability and validity and make changes to the exercise, evaluation design and instruments if needed.
8. Conduct the evaluation.
9. Analyze data and report results to the training program, curriculum developers, trainers and other stakeholders.

## Appendix J: Steps in Designing Embedded Evaluations

(Cindy Parry & Jane Berdie, Macro Eval Team Meeting April 2004)

---

### Step 1: Identify Purpose

- Criteria* Several purposes for the evaluation may be chosen. The most common include:
- providing feedback for course improvement,
  - demonstrating that training has increased the participants' skill levels (either overall in aggregate, or individual skill levels), and
  - demonstrating that participants are meeting a competency standard (again either in aggregate or individually).

Different purposes have different design and resource implications so it is important to clarify them early. For example, demonstrating that participants increased their skill levels as a result of attending training requires two parallel sets (in content and difficulty) of evaluation materials that can be used as pre and post assessments. This requires more classroom time for evaluation and more development time for the evaluation materials than a situation in which the assessment can be done once as a posttest (as when the purpose is to see if the individual or group has met a competency standard).

- Roles* Curriculum developers, training administrators, trainers, subject matter experts and other relevant stakeholders take the lead to define the desired purpose and outcomes for the evaluation. The evaluator provides information and assistance as needed to support the decision making.

### Step 2: Identify an Appropriate Competency

- Criteria* Appropriate competencies are any that deal with teaching a skill or behavior. Some examples are: testifying in court hearings, assessing family interaction patterns, using age-appropriate interviewing strategies, writing treatment plans, and using techniques for effective time management.

- Roles* Curriculum developer, trainer and evaluator jointly choose competencies. Curriculum developer and trainer take the lead.

### Step 3: Review Curriculum and Observe Training

- Criteria*
1. Does the content provide the right information and in sufficient breadth and

- depth to support development of the skill?
- 2. Does the written curriculum follow a skills teaching model that includes explanation, demonstration, practice, feedback and discussion?
- 3. Is there enough direction to the trainer to support consistent, standard delivery?
- 4. Is the curriculum delivered as written?

*Roles* Evaluator reviews and observes with the needs of the evaluation in mind. Curriculum developer, trainer and evaluator jointly determine what changes, if any, are needed.

#### **Step 4: Identify/Construct an Appropriate Exercise**

- Criteria*
1. Does it address a skill competency?
  2. Is the exercise covering the practice portion of the skills teaching model (step 3)?
  3. Is the skill adequately taught?
    - a. Does the design include the full 5 steps of the model for teaching skills?
    - b. Is enough time provided to do each step adequately?
  4. Is the task standardized to provide a common experience and basis for feedback and evaluation? (e.g., Is a common case used to develop a case plan so that the same elements might be expected to appear and scoring can be based on those expectations?)

A “no” to any of these questions does **not** mean that the task should not be considered for embedded evaluation, but all of these points will need to be addressed and some modifications made in order to use it successfully.

*Roles* Curriculum developer, trainer and evaluator jointly choose exercise and make modifications. Evaluator takes the lead on structuring the exercise to be “evaluation friendly” and trainers/developers ensure that learning objectives are still met.

#### **Step 5: Design the Embedded Evaluation**

- Criteria*
1. What information is desired and what purposes will it be used for?
 

Purpose gets at whether you want:

    - Feedback for course improvement
    - Evidence of overall course effectiveness, or
    - Evidence that individuals have mastered the skill

Purpose also determines whether you will need an:

- Individual or group response from the exercise
- Data from a sample vs. every trainee
- Anonymous responses vs. confidential responses
- Data collection pre and post training or post only

2. What procedures will be used?
  - a. Who will conduct the evaluation?
  - b. Who will evaluate performance and provide feedback?
  - c. What instruments and instructions are needed for the trainer? Trainees?
  - d. How will data be collected and forwarded for analysis?

*Roles* Evaluator takes the lead. Agency administrator(s), training administrator, curriculum developer and trainer(s) make decisions regarding purpose and information desired. The trainer may also be asked to carry out procedures for conducting the evaluation listed under Step 2.

## **Step 6: Develop Scoring Rubrics/Evaluation Instruments**

- Criteria*
1. What will adequate performance look like?
    - a. What are the dimensions (items),
    - b. What are the levels of performance for each item and how are these levels described (“anchors”)?
    - c. What is an acceptable level of performance for each item?, and
    - d. What is an acceptable overall level of performance (passing)?

*Roles* Evaluator takes the lead and designs instruments with input from the curriculum developer and trainer(s).

## **Step 7: Pilot Test the Evaluation**

- Criteria*
1. Is the evaluation being conducted as designed?
  2. Are the evaluation instruments reliable and valid?
  3. Have the persons who are scoring the instruments (e.g. trainer, a trainee taking the role of evaluator or evaluators themselves) and providing feedback to trainees been adequately trained?
  4. Has enough time been allotted to conduct the evaluation?
  5. Are trainees and the trainer(s) comfortable with the exercises and the evaluation tools and able to use them effectively?

*Roles* Evaluator takes the lead and conducts statistical analyses. Trainer(s) and trainees provide logistical and satisfaction feedback.

**Step 8: Conduct the Evaluation**

*Criteria* Is the evaluation being conducted as designed?

*Roles* As agreed to in step 5, the design phase.

**Step 9: Analyze Data and Provide Feedback**

*Criteria* 1. Have the evaluation questions been answered?  
2. Have the findings been communicated clearly and to all stakeholders?

*Roles* Evaluator produces this with stakeholders having a chance to review and comment.

# **Appendix K: Embedded Evaluation Planning Worksheet for Child Maltreatment Identification at the Skills Level of Learning**

*(Cindy Parry & Jane Berdie, Macro Eval Team Meeting April 2004)*

*\*\*\*Note: Answers in italics are based on discussion at April 2004 Macro Eval Meeting\*\*\**

---

## **Step 1: Identify Purpose/Desired Outcome of the Evaluation (Statewide)**

1. Feedback for course improvement? Yes  
*Rationale: This is a useful byproduct of almost any evaluation of training.*
  
2. Demonstrating training's role in developing skill? Likely not (If yes, requires pre post assessment)  
    \_\_\_\_\_ for participants as a whole  
    \_\_\_\_\_ for individual participants

*Rationale: The "likely not" is based on the requirements for a pre-test as well as a post test. For skills evaluation, two tests (one pre and one post) that are comparable (but not identical) need to be developed and administered. The reason is that learning can occur during a pre-test and this affects the post test. Having two comparable tests is highly labor intensive, both for development and administration (taking time from the classroom day). While eliminating the pre-test means that we can't know for sure that training is responsible for learning, a post test only design does help us to know the extent to which the skill is being acquired and that is the ultimate purpose of training.*

3. Demonstrating competency? Yes (If yes, requires post assessment only)  
    Yes for participants as a whole  
    Likely not for individual participants

*Rationale: Use of testing for individual participants has many implications that make it difficult. Generally, one would not want to make a statement about an individual's competency based on one relatively short test. Additionally, only the narrow range of focus of the skill being tested could be commented on (and there is a concern about extrapolating from this narrow set of data about the person's overall competency). Also, at this point the test would still be in the testing phase itself.*

4. Other: \_\_\_\_\_

## **Step 2: Identify Appropriate Competency/Objectives for Evaluation (Statewide)**

1. What competencies/objectives should be the focus of the evaluation? (note that some knowledge level is needed to under-gird the skill level)  
Competencies from Standardized Core:
  - a. Knowledge: *The worker will accurately differentiate between the factors that constitute abuse and/or neglect, and normative parenting styles.*
  - b. Skill: *The worker will identify behavioral characteristics of children who have been maltreated.*

Learning Objectives from Standardized Core:

yes The worker will understand the legal basis for identifying abuse and neglect in California, and understand the associated sections (A-F) of the W & I Code.

*Rationale: This knowledge level objective under-girds the skill. While assessing it is not an explicit part of the embedded evaluation exercise, this content would be taught in the module as a precursor to performance of the skill and might be tested using the item bank for knowledge tests.*

yes Given a case scenario, the worker will be able to determine whether physical abuse has occurred, according to the legal definition of abuse in California Penal Code and Welfare and Institutions Code.

*Rationale: The focus of the training and evaluation should probably be on physical abuse only since it is the clearest and has less inter-county variation in decision making. The focus would be on the assessment needed to make a determination about whether what has already happened meets the statutory requirements for "abuse." This is not to include assessment of future risk or current safety.*

Likely not Given a case scenario, the worker will be able to determine whether neglect has occurred, according to the legal definition of abuse in California Penal Code and Welfare and Institutions Code.

*Rationale: See above.*

No Given a case scenario, the worker will be able to determine whether emotional abuse has occurred, according to the legal definition of abuse in California Penal Code and Welfare and Institutions Code.

*Rationale: Too complex, rarely cited as sole allegation, and not as frequent as physical abuse*

2. Within each competency/objective, what is the key content that should be evaluated?

*The focus would be the ability to assess various information about a case, to make a decision about whether physical abuse as defined by W&I Code occurred and (possibly) to be able to identify the factors that helped in making the decision (the latter might include categories of information such as the nature of the injury, the plausibility of the various explanations of the injury, and related behavioral characteristics).*

**Step 3: Review Existing Curricula and Observe Training to Identify Potential Components for the New Common Curriculum Module (RTA and County)**

1. Is the content at the right level, relevant, current? If not, what needs to change?

2. Is the target skill adequately taught?
  - a. Does the design include the full 5 steps of the model for teaching skills?
  
  - b. Is enough time provided to do each step adequately?
  
3. Is enough direction provided to the trainer to promote complete and consistent delivery?
  
4. Is the training delivered as written? If not, what can be done to change this?

*One approach to step 3 would be that the subject matter experts on the Content subcommittee would review their own curricula in terms of content and skill level of training. This would be helpful in identifying content and training methods that could be used in the "new" module.*

#### **Step 4: Develop the Framework for the Training Module to Teach the Skill (Statewide)**

1. What are the implications for the new training module of RTAs & counties sharing information from Step 3?
  
2. What are the guidelines/parameters regarding content and delivery methods for the new training module based on the five steps of skill training?
  - a. Explain: is the relationship between the skill and knowledge made sufficiently clear?
  
  - b. Demonstrate: is the skill demonstrated (e.g., by the trainers or on videotape) so that participants see the skill done correctly?
  
  - c. Practice: are there one or more structured, standardized practice sessions?
  
  - d. Feedback: is there a structured approach for participants to receive feedback, preferably from the trainer and/or other skilled person? If participants give feedback, have they been trained to do so?
  
  - e. Discuss transfer: is there a structured approach for discussing use of the skill on the job?

## Step 5: Create Parameters for Design the Embedded Evaluation (Statewide)

1. What are some basic design parameters?:  
\_\_\_\_\_ Individual responses or group responses to the exercise

*Rationale: Individual responses allow the evaluators to link responses to information that could help explain response patterns, e.g., whether the skill is being mastered better by participants who have prior experience or education. This could help in tailoring training among the various sites. Individual responses also provide more data more quickly. More data allows for a shorter piloting process in order to collect enough feedback to finalize the evaluation instruments. More data also allows better estimates of the overall performance of the group (e.g., in a 30 person training your judgment about how well the group can perform the skill is based on thirty responses, rather than 6 as would be the case if one response was collected for each of 6 five-person groups). It is important to note that collecting responses from every individual is not the same thing as providing individual level feedback. These responses frequently are aggregated and used only to make decisions about the groups' (or a subgroups' e.g., MSWs) performance.*

\_\_\_\_\_ Data from a sample or data from every trainee

*Rationale: All trainees would participate (because the testing would be part of the training day and because part of the value of the embedded evaluation is to reinforce learning), so, it makes sense to use all data.*

\_\_\_\_\_ Anonymous responses or confidential responses (coded to prevent identification of respondent except by evaluation staff)

*Rationale: Confidentiality means that the person's name and performance are not shared but the evaluators are able to link performance to information that might help explain patterns (see above).*

\_\_\_\_\_ Data collection pre and post training or post only  
*See first page of this handout for rationale.*

2. What will the evaluation exercise be?
  - a. Extension of training activity or new activity?

*Rationale: A new scenario is easier to construct and is not "contaminated" by the scenario used in the training portion.*

- b. Slides, scenarios or combination?

*It might be useful to put together a wider range of information that would be the basis for the evaluation, e.g., slides of injuries, a written case scenario, and brief audio tapes (to be played by the person administering the exam) that would consist of snippets of interviews (worker and child, worker and parent) and meetings (worker and supervisor discussing findings). The participants would read, listen to and look at all of this information and then record responses to a series of questions on a paper test or Classroom Performance System (CPS) – the clicker system.*

*Rationale: This type of exercise might be closer to what the actual decision making process might include than simply looking at a slide of an injury.*

- c. How long to spend and how many items?

*This will depend on the final decisions made about content to include and the format of the items. A minimum of 30 minutes is likely to be needed for the evaluation portion of the skills module, but a final*

*recommendation about test length can not be made until the scope and format of the evaluation have been agreed upon.*

3. What procedures will be used?

*Logistics will need to be worked out but can be variable across locations as long as standards for administering and scoring the evaluation are consistent and consistently applied.*

a. Who will administer the evaluation?

Trainer?

RTA or county staff development person?

*Will vary by locale*

b. Who will score the participants' performance on the exercise?

Trainer?

RTA or county staff development person?

*Will vary by locale*

c. When and how will feedback be provided to participants?

*Participants would: "turn in" their tests and then the trainer would process it with them. This reinforces learning. Participants should turn in their tests (rather than take them with them) to prevent test content from circulating and future results from being inflated.*

d. What instruments and instructions are needed for the participants/trainers/test administrators?

*Trainers and (if different) test administrators need to have written material and training about the test and how to administer it and (trainers) how to facilitate the processing after the test. The participants need information about how the test will be used and how it will NOT be used (e.g., no feedback to their supervisors about individual performance). Scoring rubrics (guides) and training on their use frequently are also necessary to score the evaluations consistently and fairly (see step 6).*

e. How will data be collected and forwarded for statewide analysis?

*Depending on whether paper/pencil and/or clicker system is used, the logistics of this are TBD.*

f. What will be reported and to whom?

*Everyone gets statewide data. Regional data go only to relevant RTA/IUC. County data go only to relevant county.*

*Rationale: This reflects a previous decision by the macro-evaluation team.*

## Step 6: Develop Scoring Rubrics/Evaluation Instruments

### 1. What will adequate performance look like?

#### a. What are the dimensions (items)?

*(e.g., for Child Maltreatment)*

- *What maltreatment types will the items cover?*
- *What examples will be covered within the broad maltreatment categories (e.g., Immersion burns?, Spiral fractures?, Dirty houses?)*
- *What common conditions that are not maltreatment will be included? (e.g. Mongolian spots)*

*This will follow from decisions made about the scope of the evaluation, however it will be more detailed. We recommend the development of a test plan that specifies how many items to develop and in what areas, to help ensure that the evaluation matches the important content taught (content validity).*

#### b. What is an appropriate response format?

\_\_\_\_\_ Yes or No

\_\_\_\_\_ A rating of how likely it is to be maltreatment (e.g. on a 5 point scale)

\_\_\_\_\_ A description of an action to be taken

\_\_\_\_\_ A narrative describing their choice and rationale

*This can vary and needs to be discussed more as the training material and the test are being developed. The advantage of answers that can be quantified is that scoring is faster and less likely to vary by who is the scorer, but it is important to note that all narrative answers can be quantified as long as the anchors are clear and raters have been trained and themselves evaluated for inter-rater reliability.*

#### c. How should anchors be developed?

*Anchors are narrative descriptions of points on a scale (e.g., what constitutes a rating of 1, 2, or 3). Subject matter experts are key to establishing valid anchor descriptions. Anchors are an important aid to making scoring more uniform and consistent.*

### 2. How will acceptable performance be determined?

#### a. **For an item?** (i.e., agreement on what is an acceptable/correct answer)

*Subject matter experts are key for this.*

#### b. **For an individual?** (Needed if using a post only, criterion referenced design, even if there are no plans to report individual scores. Still becomes part of determining if class met an aggregate standard.)

*If you have more than one item, what do they have to get for their overall performance to be acceptable? 8 out of 10? 9 out of 10?*

#### c. **For the group** (Optional depending on design. Used to determine if group met competency standard with post-only criterion referenced design intended to evaluate training's effectiveness, not individual competency).

<i>Sample Competency Standards</i>		
<b>High Level</b>	<b>Beginning Level</b>	<b>Unacceptable</b>
Caseworker identifies maltreatment indicators for all or nearly all slides or scenarios (90%-100% accuracy)	Caseworker identifies maltreatment indicators for most slides or scenarios (70%-90% accuracy)	Caseworker identifies maltreatment indicators for less than 70% of the slides or scenarios.

*Three levels of decision:*

1. *Adequate performance on individual questions*
2. *Does individual pass or not pass?*
3. *Group: adequacy of the training (instead of the individual training). Is 80% of the group meeting the competency standard good enough? 90%? What's good enough as our training goal?*

These are subject matter expert and policy decisions and will need additional discussion. It is frequently helpful to make these decisions after the pilot phase of a project after the evaluation has been finalized and data on participant performance are available.