Evaluation of the California Common Core for Child Welfare Training:
Implementation Status, Results and Future Directions

Leslie W. Zeitler, LCSW, Cynthia F. Parry, Ph.D., Barrett L. Johnson, LCSW, and Jane Berdie, M.S.W.

December 2009

Executive Summary

Driven by an outcome focus, Child Welfare is evolving into a more accountable system. Federal and state measures are now used by state and local agencies in their attempts to systematically evaluate and improve their programs. Given that training of staff is often used as a key strategy or intervention to change practices and influence outcomes, it is crucial that training resources be evaluated in order to assure they are effective in achieving the intended knowledge, skill, and program impact. Being able to determine the efficacy and utility of training is important not only to the California Department of Social Services and the federal government (both of which expend considerable financial support toward training efforts), but also to the specific counties which hire and retain child welfare workers and supervisors.

Via the Macro Evaluation Subcommittee (a subcommittee of STEC, the Statewide Training & Education Committee), California began to strategically plan for evaluating child welfare in-service training in 2002. Implementation of the evaluation began in 2004 with the adoption of the Framework for Training Evaluation for the Common Core Curricula[1], developed as part of the Child and Family Services Review (CFSR) process by CalSWEC and its partners, the Regional Training Academies (RTAs), the Inter-University Consortium (IUC), and the counties. The systematic review of child welfare training is a considerable effort because of the complex nature of training delivery in California and the highly technical training content.

Evaluating data at each level progressively (working from Level 1 up through Level 7) helps to create a "chain of evidence" to determine the extent to which training has had an impact on trainees. The table below outlines the levels of evaluation, the results of data analysis, and the importance to counties of information gathered at each level.

Level of Evaluation	Results	Importance to Counties
Level 1: Tracking Training— Tracking attendance/ demographics— Evaluate who attended.	Comprehensive demographic data have been captured for nearly all new child welfare social workers and supervisors (3,692 trainees as of 12/31/08) since formal evaluations began in 2005. These data assist in assuring that test materials are fair and valid, and provide an ongoing picture of the demographic profile of new child welfare professionals. Completion of the Common Core is tracked systematically by the RTAs/IUC; upon completion, data are shared with the counties.	Statewide training regulations adopted in July 2008 require that counties track completion of the Common Core for child welfare social workers and supervisors.
Level 2: Formative evaluation of training courses (course level)— Evaluate what is intended to be taught in the content, how it is	As part of ongoing revision processes, systematic collection and analyses of data regarding training content and delivery have resulted in improvements to the Common Core.	As part of the evaluation at this level, counties and other stakeholders assist in the revision of the Common Core, assuring that it is relevant to practice and imparts key knowledge that child welfare staff need to complete their duties.

[1] Parry, C., & Berdie, J. (2004). *Training Evaluation Framework Report*. Berkeley, CA: California Social Work Education Center.

taught, and how it was developed and revised.		
Level 3: Trainee Satisfaction—Evaluate how satisfied the participants were with the training experience and their opinions about its usefulness.	This level of evaluation is completed at a regional level, statewide.	A major benefit to county staff of trainee satisfaction evaluations is that trainees are encouraged to provide information that allows for adjustments to the breadth and/or depth of training content—which also can affect applicability to the job.
Level 4: Trainee knowledge acquisition—Evaluate knowledge gained as a result of training.	There were three topics in which knowledge tests were administered and post-training (hereinafter referred to as pre-tests and post-tests): *Child & Youth Development in a Child Welfare Context, Family Engagement in Case Planning & Case Management,* and *Placement & Permanency*. Four years of data indicate that overall, trainees (new child welfare workers) improved at a statistically significant level in their scores from pre- to post-test. In some years and for some curricula, Title IV-E trainees achieved significantly higher scores than non-Title IV-E trainees at pre-test and at post-test. However, they also achieved statistically significant gains from pre-test to post-test, indicating that they gained knowledge as a result of the training. For the topic for which a knowledge post-test only is administered (*Critical Thinking in Child Welfare Assessment: Safety, Risk & Protective Capacity)*: Although no formal standard has been established that serves as a yardstick of mastery, the data indicate that trainees leave the classroom with a substantial level of knowledge related to course learning objectives.	The original strategic plan focused on Levels 4 and 5. County management personnel want to know that their staff are adequately prepared to meet the demands of child welfare work with basic knowledge (e.g., including knowledge related to identifying the various forms of child maltreatment and applying relevant state/federal laws to direct child welfare work).
Level 5: Skills acquisition by the trainee as demonstrated in the classroom—Evaluate skills gained as a result of training.	For topics in which skill is assessed in the classroom (e.g., embedded evaluation[2]) pertaining to identification of physical abuse and sexual abuse (*Child Maltreatment Identification, Parts 1 and 2):* At least 80% (and in most years 90% or more) of new child welfare workers made 3 out of 4 correct decisions when asked to indicate whether or not child maltreatment occurred in a given case scenario.	The original strategic plan focused on Levels 4 and 5. County management personnel want to know that their staff are adequately prepared to meet the demands of child welfare work with a basic set of skills (e.g., including skills related to identifying the various forms of child maltreatment and assessing safety and risk).

[2] The term "embedded evaluation" refers to an evaluation or assessment that occurs in the classroom setting that is simultaneously utilized as a training tool (reinforcement of learning).

Level 6: Transfer of Learning (TOL) (use of knowledge and skill on the job)—Evaluation is used to measure the extent to which a participant can transfer the learning from the classroom to the work setting.	Evaluation at this level has not yet begun on a statewide basis. However, regional studies have been completed and are discussed further in this report.	County staff may be particularly interested in knowing how to support TOL at the county level and determining which TOL strategies are most effective. This helps determine allocation of funding, time for training, and supervisory staff follow-up. It is imperative that training in the classroom translates back to practice in an effective manner; otherwise, the resources allocated to training could be better spent elsewhere.
Level 7: Agency/client outcomes—Evaluation is used to measure the degree to which training affects achievement of specific agency goals and/or client outcomes.	While we have yet to undertake projects of this nature due to current resource issues and the need to develop the building blocks at the lower levels in a rigorous manner (as part of developing a chain of evidence), it is the eventual goal of the child welfare training evaluation system in California to link training interventions to outcomes for children and families served by CWS.	Understanding the impact of training on outcomes in our counties is especially important in a budgetary climate when many child welfare services are being cut.

The evaluation system has produced data worthy of examination. As the results indicate, California has achieved significant progress in systematically evaluating child welfare training, and has built an infrastructure for training evaluation that creates opportunities for building upon existing findings. Moving forward, the statewide child welfare training system plans for the following activities as part of the next three-year strategic plan:

- Complete demographic analyses of Title IV-E trainee test data.
- Develop, review, and revise knowledge tests for applicable curricula.
- Pilot the use of a neglect scenario as part of an embedded evaluation.
- Revise and implement the embedded evaluation in the *Casework Supervision* module of the Common Core for Supervisors.
- Pilot embedded evaluation for Structured Decision-Making™ (SDM)/*Critical Thinking* training that includes SDM Hotline tools.
- Pilot attitude/values evaluations related to the *Child Maltreatment Identification, Parts 1 and 2* curricula.
- Develop a model for trainer evaluation.
- Launch statewide quality assurance efforts for child welfare worker training.
- Initiate and complete a feasibility study with respect to a Transfer of Learning evaluation.
- Initiate and complete a feasibility study with respect to an outcomes evaluation.

Activities may be amended as necessary.

Acknowledgments

California's efforts to support an effective statewide training and evaluation system in child welfare—and the Framework for Training Evaluation used to guide this system—is the result of the invaluable work and guidance of a great many people throughout the child welfare system in California and across the country. It would be impossible to list all of the individuals who contributed in some way, but some groups of people will be acknowledged here.

The Statewide Training and Education Committee (STEC) provides overall guidance for the development of the curricula for which trainings are evaluated. Convened by the California Social Work Education Center (CalSWEC) and the California Department of Social Services (CDSS), STEC has a wide membership that includes Regional Training Academy (RTA) representatives (the Bay Area Academy, the Central California Training Academy, the Inter-University Consortium in Los Angeles (IUC), the Northern California Training Academy, and the Public Child Welfare Training Academy—Southern Region); county representatives; university-based Title IV-E Project Coordinators; and other key stakeholders.

A subcommittee of STEC, the Macro Evaluation Subcommittee provided oversight and approval for the development and implementation of the statewide training evaluation process. The consistent participation of the Regional Training Academy/Inter-University Consortium staff in training evaluation efforts has greatly increased what we know about trainees when they leave the training room in terms of knowledge and in some cases, skill. The Macro Evaluation Subcommittee also provided feedback on this report.

National training evaluation consultants support our ongoing efforts, utilizing the breadth and the depth of their expertise in child welfare training evaluation, and help us apply evaluation principles and strategies to a complex training system. Along the way, many other people provided their insight and hard work, attending pilots of the trainings, reviewing evaluation materials, or providing other assistance.

Without the dedicated work of our partners, CalSWEC and CDSS would not be able to provide a perspective on what is happening in training rooms around the state; nor would we be able to establish the initial building blocks toward the eventual evaluation at the level of outcomes. We wish to express our sincere appreciation to all of our partners for their dedication in working to improve child welfare services through the lens of training and evaluation. Without their contributions, we would not have had the information necessary to evaluate training throughout California.

Table of Contents

Executive Summary ..... *i*
Acknowledgments ..... *iv*
I. Introduction *1*
II. Background and Context *2*
III. Implementation Status and Results *7*
Level 1a and 1b: Tracking Attendance and Participant Demographics 7
Level 2: Course Evaluation 10
Level 3: Satisfaction/Opinion 13
Level 4: Knowledge 14
Level 5: Skills 21
Level 6: Transfer of Learning 25
Level 7: Agency/Client Outcomes 30
IV. Analysis of Progress and Findings 33
V. Future Plans for Evaluation *35*
VI. Bibliography *39*
VII. Appendices *40*
Appendix A: Levels of Training Evaluation *41*
Appendix B: The Chain of Evidence *43*
Appendix C: CDSS Common Framework for Assessing Effectiveness of Training:
A Strategic Planning Grid Sample Framework for Common Core *45*
Appendix D: Standardized ID Code Assignment Instructions *49*
Appendix E: Demographic Survey (revised March 2009) *50*
Appendix F: Sample Delta Plus Tool (Course Level Evaluation) *52*
Appendix G: Sample Global Evaluation Tool (Course Level Evaluation) *56*
Appendix H: Proposed Timelines for Revised Curricula/Evaluations *60*
Appendix I: Sample Rasch Analysis of Test Items (Validation process) *61*
Appendix J: Item Piloting Update (March 2009) *65*
Appendix K: Protocol for Revision of Problematic Knowledge Test Items *66*
Appendix L: Summary of Knowledge Test Performance (one curriculum) *69*
Appendix M: Summary of Test Performance by Title IV-E/non-Title IV-E Status *72*
Appendix N: Sample Summary of Performance on Embedded Evaluations *82*
Appendix O: Strategic Plan for Child Welfare Training Evaluation 2009-2012, Summary Matrix *93*
Appendix P: Strategic Plan for Child Welfare Training Evaluation 2009-2012, Gantt Chart *102*

I. Introduction

Child welfare is evolving into a more accountable system, driven by outcomes. Federal and state measures are now used by state and local agencies in their attempts to systematically evaluate and improve their programs. These agencies often use training of staff as a key strategy or intervention to change practice and influence outcomes. In response, child welfare training systems around the country have grown in scope and sophistication.

Until recently, however, most child welfare training systems relied on little more than trainee satisfaction to assess training effectiveness. Given the resources expended on training, a more systematic approach to training evaluation is called for—a single approach that evaluates the impact of training at multiple levels and provides data on training effectiveness. Such an approach requires extensive planning and strategizing even in a relatively simple, centralized training system. A complex, decentralized system such as California's—which is county-administered with state oversight and regional training entities—presents an even greater challenge.

California's training system began to strategically plan for this type of evaluation in 2002. Implementation of the evaluation began in 2004 with the adoption of the Framework for Training Evaluation for the Common Core Curricula, developed by CalSWEC as part of the Child and Family Services Review process. Since that time, significant progress has been made in implementing the Framework, with great effort by CalSWEC, the California Department of Social Services (CDSS), the Regional Training Academies (RTAs), the Inter-University Consortium in Los Angeles (IUC), and California's counties.

This report summarizes that progress and provides guidance for developing a strategic plan for the next three-year period. It includes:

- background and contextual information, including the purpose and conceptual basis for the Framework,
- implementation status and results for each level of the Framework,
- analysis of progress and findings, and
- future directions for evaluation.

II. Background and Context

The original Framework for Training Evaluation was developed by CalSWEC, with guidance and advice from the Macro Evaluation Subcommittee of California's Statewide Training and Education Committee (STEC).[3] Development and implementation of the Framework was required as part of California's 2002 Program Improvement Plan (PIP). CalSWEC consultants Jane Berdie, M.S.W., and Cynthia Parry, Ph.D., were instrumental in facilitating the development and implementation of the Framework.

Purpose of the Framework for Training Evaluation

The initial Framework's purpose was to provide a structure and organization for making specific decisions about which evaluation projects to pursue and why. It allowed for implementation of various training evaluation projects incrementally, building towards an array of projects over various levels of evaluation.

Conceptual Basis for the Framework for Training Evaluation

Levels of Evaluation

Training is traditionally evaluated primarily by assessing trainee reactions, such as their satisfaction with the training and their opinions about its usefulness on the job. Informally, it is often evaluated by the trainers and sometimes by an advisory group who view written material and observe delivery to assess the relevance of content and the degree to which the methods of delivery and content hold the trainees' interest. Occasionally knowledge is tested before (pre-test) and after training (post-test) on a given topic.

More rigorous approaches to training evaluation identify levels of evaluation, determine which are most useful, and design procedures and data collection instruments to evaluate at each level (Kirkpatrick, 1959; Parry & Berdie, 1999; Parry, Berdie, & Johnson, 2004). California's Framework for Training Evaluation utilizes levels adapted from a model developed by the American Humane Association (AHA). In this model, lower levels of evaluation (e.g., trainee satisfaction, trainee opinion, etc.) are more closely associated with an actual training intervention than are higher levels of evaluation (e.g., skill transfer, agency impact, client outcomes, etc.). This is because factors other than the training itself are increasingly likely to interfere with identifying and quantifying the cause-effect relationship between the training and the evaluation findings at higher levels of evaluation.

For example, when a course curriculum and delivery are evaluated (a formative evaluation), usually very little "interference" is present. Review of written materials and training delivery by knowledgeable, trained observers using standardized tools should produce evaluation information directly related to the actual training event. Similarly, trainee satisfaction can be fairly accurately measured using carefully designed written evaluation tools. Subsequent levels

[3] The Macro Evaluation Subcommittee began meeting in 2002 and consisted of representatives of the five Regional Training Academies/Inter-University Consortium (RTA/IUC), CalSWEC, several county agency training staff, and CDSS.

of evaluation are increasingly affected by intervening variables related to the trainee, the training design and delivery, and the organizational context in which the trainee operates.

Findings about knowledge acquisition are fairly easily measured using pre- and post-tests, but may be affected by such factors as trainee motivational differences and education levels. Similarly, observations of trainees' skills in the classroom may allow demonstration and evaluation of skill acquisition under controlled conditions, but supervisor support may affect transfer of those skills to the workplace. Factors entirely outside of training, such as caseload size and agency policies, may greatly affect the impact of training on client outcomes. In order to have a meaningful evaluation of the impact of training, one must strategically evaluate and analyze the results at multiple levels in order to make inferences about impact on outcomes for the agency and for clients.

For more information on AHA's model of training evaluation, refer to Appendix A.

The Chain of Evidence

Utilizing such levels of evaluation as part of the planning process allows training systems to establish a "chain of evidence" that logically makes connections between the training and agency and client outcomes. 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 conduct the types of studies that would establish a direct cause-and-effect relationship between training and a change in the trainee'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 in a randomized control trial.

When faced with a situation where quasi-experimental designs are the best alternative, one can demonstrate that training plays a part in producing positive outcomes by showing a progression of changes from training through transfer and outcomes for the agency and client. This requires 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, one must also consider whether or not a particular evaluation should collect information at levels lower than the level of primary interest. *More information on the Chain of Evidence can be found in Appendix B.*

Design of the Framework and the Evaluation Projects within the Framework

Decisions about which training evaluation projects to implement are generally made within the context of answering basic questions about training effectiveness:

1. Are staff learning from training?
2. If so, how much are they learning?
3. What are they learning/not learning?
4. What are they applying/not applying on the job?

5. What factors (e.g., about the trainees, the curricula, the delivery of training, the work environment where staff then implement learning) help to explain patterns of learning and use of learning on the job?
6. What do evaluation findings suggest about changes that would make the training/ professional development system more effective?

Levels

In order to answer these questions, California's Framework was organized by multiple levels of evaluation. This multi-level system helps to conceptualize evaluation in relation to the basic questions and to categorize specific evaluation projects and link them to one another.

Seven levels were delineated:

Level 1	Tracking attendance (subsequently, level 1b—Trainee Demographics was added)
Level 2	Formative evaluation of the course (curriculum content and delivery methods)
Level 3	Satisfaction and opinion of the trainee
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)

The initial Framework report included information on the following for each level:

- 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, i.e., 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?

For a summary of the original Framework Report, including some updated information on implementation, see Appendix C: CDSS Common Framework for Assessing Effectiveness of Training: A Strategic Planning Grid.

Implementation of the Initial Framework

As each evaluation project was implemented , a rigorous design was developed and executed to ensure that findings are based on valid and reliable data—and thus can help to answer the basic questions about training effectiveness. Evaluation instruments and test items were developed and analyzed to ensure that they accurately reflect important curriculum content and collect all desired information. For example:

- The four knowledge tests associated with Level 4 each have entailed a multi-year process of item development and testing in order to form a large bank of multiple-choice test questions from which tests can be drawn;

- The three "embedded evaluations" (in-class skills assessments) associated with Level 5 have undergone several revisions as a result of trainee performance data that suggested confusion with curricula, assessment cases, and instruments;
- The Demographic Form associated with Level 1b has been modified multiple times in order to more accurately identify options within certain demographic variables. Accurate demographic data are extremely helpful in understanding findings from the other levels of evaluation.

Benefits of Implementing the Framework for Training Evaluation

Using the Framework to make decisions about and develop a multi-level evaluation plan for child welfare training has provided numerous benefits. Among other things, the Framework:

- Supports both planning of training evaluation and review of progress: It provides a concise format for reporting on evaluation projects as well as overall progress.
- Helps to conceptualize and organize multi-level evaluation, which in turn increases confidence in the findings: For example, evaluation data from knowledge tests, in-class skills assessment, *and* on the job assessments provide a much stronger picture of learning than does any one level alone.
- Via the system of levels of evaluation, helps to specify exactly what is being evaluated in any given training evaluation project by linking the project to one of the levels. Thus an in-class assessment of knowledge is differentiated from one that evaluates skill.
- Supports rigorous evaluation designs by requiring brief descriptions of the design and narrative about decisions, resources, and timeframes.
- Focuses on the importance of higher levels of evaluation (transfer of learning and agency and client outcomes) in understanding the effects of training: It helps focus planning discussions on how to achieve greater understanding of whether learning is occurring by focusing on the chain of evidence as well as the individual levels of evaluation.
- Provides evaluation data from multiple levels that help inform decisions about how to systematically improve training (including curriculum, delivery of training, and trainer development) across Californa's large and varied system (involving the RTAs/IUC, CalSWEC, the state, and the counties). For example, findings from training evaluation have already helped curriculum writers target specific areas for revision. Regional training administrators have been able to support trainers in making improvements in delivery of certain sections of curricula. The need for standardization of curricula and delivery (from trainer to trainer and region to region) is now clearer. Similar benefits accrue for training-related activities (such as mentoring and other transfer of learning supports).
- Provides data about specific trainee cohorts that can be used to make decisions about whether and how to target training for those cohorts, e.g., Title IV-E students.
- Helps to clarify what needs to be in place throughout the child welfare system for evaluation to be effective, e.g., a stable curriculum based on an accepted and well-articulated model of practice that is adhered to in actual service delivery and is supported by written procedural requirements and policy.

- Has led to the development of large multi-year databases that can be mined for research and that will help California and others in the child welfare training field better understand the effects of training and improve the training system.
- Supports local evaluation initiatives by providing a statewide comparative context, instruments, item analysis, and other technical and analytic support.

Shortcomings of the Framework for Training Evaluation

Despite the many benefits of implementing the Framework, some challenges have been identified. These include:

- Regional reports 6 to 12 months after test administration compromise their utility for counties and RTAs/IUC. The point-in-time analysis supports the state's response to the federal government's program improvement plan, but is problematic when attempting to respond quickly to improve training.
- The scope of evaluation by design is intended to be programmatic and does not address the feasibility of evaluating either individual performance or support for supervisors.

III. Implementation Status and Results

The status of the implementation and relevant results for each level are summarized below. Overall analysis of progress and results will be included in subsequent sections.

Level 1a and 1b: Tracking Attendance and Participant Demographics

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 necessary to ensure that new caseworkers are being exposed to training on all of the competencies needed to do their jobs, and that such tracking occurs consistently across the state. At the time of the initial Framework report, California had no systematic method of tracking completion of training. Child welfare workers generally completed Common Core training at their regional training academy, but the state had no requirements for initial or ongoing training. While the RTAs and IUC (and some counties) tracked training attendance, and several counties required completion, the state did not require proof that each new hire completed Common Core training.

Implementation status

As part of the 2002 PIP, STEC drafted recommendations for CDSS for training requirements, which formed the basis of new training regulations that were adopted effective July 1, 2008. The regulations require all newly hired or promoted child welfare workers to complete the entire California Common Core during their first two years on the job. Training on six specific topic areas must be completed prior to one year on the job. Newly hired or promoted supervisors must also complete the Common Core for Supervisors. After completion of the Common Core, all child welfare staff must complete 40 hours of ongoing training every two years. The regulations require counties to track completion of both initial Common Core training and ongoing training, and report to the state the proportion of their workforce that has met the requirements.

In addition, the Macro Evaluation Subcommittee developed and implemented a common demographics form that is administered to each trainee and linked to their evaluation test data via a self-generated unique identifier. Demographics are aggregated and reported statewide and by region. Demographics are also used in analysis of two other levels of evaluation, e.g., knowledge and skills (embedded evaluations[4]). *A copy of the current ID Code Assignment Instructions and the most recent Demographics form are located in Appendices D and E, respectively.*

[4] The term "embedded evaluation" refers to an evaluation or assessment that occurs in the classroom setting that is simultaneously utilized as a training tool (reinforcement of learning).

Resources expended
CalSWEC, CDSS, the counties, and the RTAs/IUC all expended resources on this level. CalSWEC developed databases, and entered and analyzed demographic data for 3,692 trainees since 2005 (through 12/31/08). CalSWEC also assisted CDSS to train county staff on the training regulations, and sponsored a Tracking Training Forum on February 20, 2008.

Counties invested personnel time and systems to track completion of training by their employees. The RTAs/IUC conformed their existing tracking systems to allow them to provide counties in their respective regions the data on completion, and expended personnel time in collecting demographic data for each Common Core trainee. Cost savings can be achieved when employees transfer from county to county, because these employees will not have to repeat the Common Core training for new child welfare workers.

Summary of Results
As a result of tracking completion and demographic data, relatively comprehensive demographic data have been available for all newly hired child welfare workers since fall 2005. These data are linked to performance on knowledge and skills tests, and also are used to assist in validating knowledge items (see Level 4).

Because reporting statewide demographic data can mask significant regional variation, aggregating most frequent responses to the demographic survey items statewide provides a composite set of trainee data in calendar year 2008. Among all trainees, on average the new worker:

- Was female (82.7%);
- Was between 26 and 35 years of age (46.5%);
- Was Hispanic/Latino (30.0%) or Caucasian (35.3%);
- Had an MSW (38.4%) or a BA/BS (30.8%);
- Had not participated in a Title IV-E educational stipend program (68.2%);
- Had more than 6 months of previous Child Welfare experience (53.9%);
- Had been in his or her current position less than 6 months (79.8%);
- Did not hold a current license as a mental health practitioner (95.1%);
- Spoke English as a first language (74.6%);
- Did not yet carry a caseload (65.8%);
- Was excited about attending Core training (71.3%);
- Was not concerned about time away from the office while attending training (76.7%);
- Had heard that Core training was valuable (65.9%);
- Had discussed training needs with a supervisor or mentor (52.7%);
- Had specific clients in mind with whom to use what was learned in training (66.0%).

Significant trends were noted from the first full year of data collection (2006) to 2008 in the areas of education, experience, race/ ethnicity, ESL status, age, caseload, and trainee readiness for transfer of learning.

- Education: In the area of education, there was a general trend toward more trainees in caseworker Common Core classes reporting higher educational levels and more professional social work preparation. Specifically, higher percentages of trainees reported:
    - Having master's-level (MA/MS) degrees (from 16.5% to 18%)
    - Having MSWs (from 31.4% to 38.4%), and
    - Participating in Title IV-E educational stipend programs (from 25.5% to 31.8%),
- Timing of training: Trainees were more likely to enter Common Core training within their first few months on the job. Percentages reporting that they had been in their current position for less than 6 months rose from 60.9% to 79.8%, and fewer trainees reported being in their current positions for more than 1 year (decreases of 5.1% for 1–2 years, 7.3% for 3–5 years, and 7.4% for over 5 years). In counties such as Los Angeles, all new CSWs are required to complete Common Core training prior to their office assignment.
- Age of trainee: Trainees were more likely to be somewhat older and to report having more previous child welfare experience. Specifically:
    - The percentage having 6 months or more of child welfare experience prior to their current position rose from 33.7% to 53.9%;
    - The percentage of trainees in the 36-to-45-year-old group increased by 2.4% and the 46-and-older group by 3.2%;
    - Percentages decreased for trainees in the 25-and-younger group by 1.7% and the 26-to-35-year-old group by 4.1%.
- Caseload: Trainees were more likely to report carrying a caseload. Percentages rose from 27.4% to 34.2%. Caseload size also seems to be increasing. Percentages reporting caseloads of 0 to 10 fell 17.4%, while increases were reported for caseloads of 11–20 (8%) and 21–30 (6.6%). This statewide trend is offset by counties such as Los Angeles, where Common Core training is pre-caseload.
- Race/Ethnicity: The two most frequently reported categories of race/ethnicity continued to be Hispanic/Latino and Caucasian. However, the percentage of trainees reporting Caucasian race/ethnicity rose from 30.7% to 35.3%, and the percentage of Hispanic/Latino trainees decreased from 38.5% to 30%. Regions and counties differed in their racial/ethnic profiles. For example, Los Angeles has a much more diverse set of hires, with Caucasian trainees representing 20% of all new trainees, African-Americans representing 28%, Hispanic/Latino representing 38%, and Asian/Pacific Islanders comprising 9%.
- ESL: Fewer trainees reported English as a second language (29.7% in 2006 compared to 25.4% in 2008).
- Potential for transfer of learning: There were no changes on questions related to the potential for transfer of learning with one exception. The percentage of trainees who indicated that they had clients in mind with whom they could use the information learned in training rose from 53.6% to 66.0%.

Assessment of Progress
In the California child welfare system, input from stakeholders (including members of the decentralized, county-based system) is required to craft workable policies. As a result, it was a significant achievement for the state to develop, adopt and implement statewide regulations for tracking the participation of new child welfare workers and new child welfare supervisors in training. Tracking training, however, is more complicated than it initially appears. Though not universal, some counties without existing electronic training data systems need a considerable amount of explicit instruction on how to count the trainees, in addition to ongoing support to track data accurately over time.

Implementation of the demographic form has created an ongoing data set with comprehensive information about the statewide workforce. The data can be sorted by county and region and presents great opportunities to learn more about the workforce and how it evolves over time as well as how various trainee demographics may be associated with learning.

Future Directions
The training regulations were implemented in July 2008; tracking systems will likely evolve and become more sophisticated in coming years. As counties become responsible for meeting ongoing training requirements, demand for advanced and ongoing training will likely increase. Counties and RTAs/IUC must also assure that demand for training does not exceed the availability of training when ongoing training deadlines approach. For example, if large counties set policies that mandate one due date for the completion of the 40 hours for all of their staff, training demand close to that deadline will inevitably exceed supply. Development of alternative methods of delivery (i.e., other than classroom training) may help to alleviate this problem.

With respect to tasks that are part of the next strategic plan, demographic profiles and related analyses of line worker Common Core test data will continue. In addition, demographic profiles and related analyses of Common Core for Supervisors test data will commence.

Analyses of Title IV-E trainee test data will also be completed as part of this next strategic plan. Such analyses include comparisons of Title IV-E status and time on the job and comparisons of Title IV-E training evaluation data with Title IV-E workforce/career path data. Additional questions to be answered include: In which content areas are Title IV-E trainees improving? In which content areas are Title IV-E trainees NOT improving? In which content areas do Title IV-E trainees come in with more knowledge? This analysis will help determine if more customized Common Core content would benefit Title IV-E students.

Level 2: Course Evaluation

While course-level evaluation initially appears straightforward, California's decentralized training system presented many challenges in terms of content development and

standardization. Prior to the implementation of the initial Framework, Common Core training content varied greatly by region. Although previous attempts were made to develop a fully standardized curriculum statewide, none had been implemented. Before evaluating curriculum content at this level, the state needed to develop a suitable system of standardization for statewide Common Core curriculum.

Scope and Description

As part of California's PIP (and simultaneously with the development of the Framework for Training Evaluation), STEC identified seven areas of the Common Core as priorities for curriculum content standardization and evaluation (hereafter called the "Big 7"):

- *Framework for Child Welfare Practice in California* (like the other areas, this has standard content, but is the only one not evaluated)
- *Child & Youth Development in a Child Welfare Context*
- *Critical Thinking in Child Welfare Assessment: Safety, Risk, and Protective Capacity*
- *Child Maltreatment Identification, Part 1: Neglect, Physical Abuse, and Emotional Abuse*
- *Child Maltreatment Identification, Part 2: Sexual Abuse and Exploitation*
- *Family Engagement in Case Planning and Case Management*
- *Placement and Permanency*

STEC also identified six content areas of the Common Core for Supervisors for standardized content:

- *Casework Supervision*
- *Child Welfare Policy and Practice for Supervisors*
- *Evidence-Based Practice*
- *Fiscal Essentials*
- *Educational Supervision*
- *Managing for Results*

The Casework Supervision module of the Common Core for Supervisors includes an evaluation component.

Systematic development and revision of the Big 7 content areas occurs on an ongoing basis. Under the direction of the Content Development Oversight Group (CDOG), a subcommittee of STEC, each of the Big 7 originally was the responsibility of one or more of the RTAs/IUC or CalSWEC. In each case, the lead entity reviewed and updated 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. Subsequently, revision and development were centralized under STEC, with CalSWEC coordinating the activities.

At the course level, evaluation activities focus on the further development of curriculum content, guided by a process of formative evaluation and quality assurance. Course-level evaluation data are used to update the curriculum systematically for each major revision.

Implementation Status
CDOG developed specific protocols and forms for construction and review of curricula, and to assist with review (observation) of actual training delivery prior to March 2005. CalSWEC administered the collection of and aggregated the data from these forms during pilot and curriculum revision processes, beginning in 2005. The review forms have been used in piloting and subsequent revisions of every topic of Common Core for line workers and supervisors.

In 2007, STEC recommended the centralization of curricula revision at CalSWEC. Under the direction of CDOG and STEC, CalSWEC coordinates the collection and analysis of course-level data, and develops revised curricula for the Common Core.

Sample course level evaluation tools are included as Appendices F (Delta Plus Tool) and G (Global Evaluation Tool).

Resources Expended
Initial standardization of curriculum content required resources from all training entities in the state. Each lead organization contributed personnel and/or consultants to form a workgroup, review and revise competencies and learning objectives, review literature as needed, develop new curricula as needed, adhere to CDOG's decisions and protocols regarding quality assurance and curriculum format, and finalize the curriculum content. Each lead entity was asked to participate in CDOG meetings to guide and track progress.

Under the revised plan, CalSWEC incurs more costs associated with curriculum development, but each region expends significant staff time to participate actively in CDOG's curriculum oversight activities.

Summary of Results
There are no formal results for this level of evaluation other than the systematic use of feedback in curriculum revisions. The standardized curricula itself is the result of the process.

Assessment of Progress
The system of course-level evaluation is invaluable to the curriculum development and revision process. The results of these course-level evaluations are highly standardized learning objectives and content for all Common Core training. All Common Core curricula have been piloted, revised, and re-launched, and are now in a regular cycle of revision.

Several lessons were learned from the implementation of course-level evaluations. Simultaneous development of multiple curricula and implementation of the respective course-level evaluations significantly overtaxes necessary personnel. Also, curriculum content should be stable before implementing higher levels of evaluation connected with it. For best results, evaluation and curriculum activities must be carefully coordinated.

See Appendix H for the combined curriculum and evaluation development/revision timetable submitted to STEC in December 2007.

Future Directions

The system of course-level evaluations is very detail-oriented and mixes feedback on training delivery and training content, which is difficult to integrate. Analysis suggested that the course-level evaluations of training should separate assessment of content from assessment of training delivery in order to better integrate constructive feedback in subsequent versions of the curricula. This involves content experts in the field reviewing the curriculum for accuracy, research-base, etc., and experienced trainers reviewing the curriculum for delivery, integration of adult-learning methods, etc. CDOG plans to separate these functions more distinctly in future years.

As part of the next strategic plan, formative evaluations for observers will be divided into assessments of content and assessments of delivery. This includes development of formative evaluation materials for a new statewide venture: the e-learning platform.

Level 3: Satisfaction/Opinion

Scope and Description

In formulating the initial Framework, CalSWEC and the Macro Evaluation Subcommittee determined that data at this level did not require standardization, since there were no plans to link satisfaction to performance on knowledge or skill assessments. The RTAs/IUC currently use forms to collect participant feedback on the quality of training. In addition, the IUC's evaluation system includes electronic, online data entry of participant reactions and summary reporting.

Implementation Status

These evaluations are in place and currently being conducted at the RTA/IUC and county level.

Resources Expended

Since the RTAs/IUC already collected this data, no extra resources were expended as a result of the Framework's implementation.

Summary of Results

There are no results, given that statewide evaluation at this level is not part of the Framework.

Assessment of Progress

The RTAs/IUC continue to collect satisfaction-level evaluation data.

Future Directions
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.) Research suggests that perceptions of training utility affect trainees' transfer of training to the job (Alliger et al., 1997).

Level 4: Knowledge

Scope and Description
The initial Framework called for development and implementation of knowledge tests for the four content areas with standardized content, but with some variation in delivery of the training:

- *Child and Youth Development in a Child Welfare Context*
- *Critical Thinking in Child Welfare Assessment: Safety, Risk & Protective Capacity*
- *Family Engagement in Case Planning and Case Management*
- *Placement and Permanency*

The purpose of the knowledge testing component of the Framework was to provide feedback on gains in trainee knowledge relative to key course content to be used for course improvement. Aggregate results from the testing serve as evidence of the effectiveness of training in helping trainees acquire job-related knowledge.

Using standard protocols, the RTAs/IUC and any county providing Common Core training administered pre- and post-training multiple-choice tests for the three content areas that involved more than one day of training. In order to conserve training time, only one training curriculum incorporated the use of a multiple-choice test only at the end of training: *Critical Thinking in Child Welfare Assessment*, which is one day long.

For each of these four content areas, the tests include 25–30 multiple choice test questions that tap knowledge covered in the curriculum as well as application of concepts to short case-based scenarios. Specific items for each test version cover the full spectrum of learning objectives.

CalSWEC acquired approval for the evaluation procedures from UC Berkeley's Committee for the Protection of Human Subjects (CPHS). Participants are informed of the purpose of the evaluation, confidentiality procedures, and how the results will be reported and used. Trainers have written instructions and/or training on how to administer and debrief evaluations and monitor the evaluation process. With the exception of Los Angeles, evaluation results are not shared on an individual level with counties.[5] For the security of the multiple-choice test questions ("items") used in the tests, participants turn in their tests before leaving the classroom

[5]. Separate consent forms are used, since the information about the employee's performance may impact the impression of the employee by their supervisor.

to avoid a loss of item validity if they are circulated. In Los Angeles, statewide and county-developed test items are scored and released to trainees and supervisors as a Common Core evaluation transcript. Using supervisors as subject matter experts, Los Angeles developed "minimal competency scores" for some training evaluation. The transcripts support supervisors in supervision, training, and mentoring.

Test Item development

Teams of content and curriculum experts develop and refine the multiple–choice test questions (hereafter referred to as "items") for initial use in the tests. Test items are mapped to learning objectives. This assures that tests contain items for key course content. CalSWEC develops tests centrally, and test results are sent to CalSWEC for data entry and analysis. Upon validation *(see below)*, items are included in a database of test items[6] that can be used interchangeably for future tests. Specialized item-banking software will be used to generate tests once the bank is developed.

Test item validation

Test items are considered valid if they have a clear relationship to important curriculum content and measure what they are intended to measure: the trainees' knowledge of key content. Item analysis identifies items that are not functioning as valid measures of knowledge due to either deficiencies in the items' structure or their relationship to curriculum content. Scores on the CalSWEC knowledge tests are computed based only on those items that meet criteria for a well-functioning and content-valid item.

Test items used in the CalSWEC tests have been validated through several means:

- by linking them to learning objectives, course content, and research as mentioned previously,
- by editorial review to ensure clarity of the item, and
- statistically, using Rasch modeling with data from tests taken by trainees to assess the item's difficulty and the ability of the item to differentiate test-takers who perform well on the tests as a whole from those who perform less well.

Using the demographic data linked by a unique identifier code, CalSWEC analyzes the validity of test items to determine if particular demographic characteristics (such as race and ethnicity) were associated with a correct or incorrect answer. Test-takers who perform well on the tests as a whole should perform well on individual items, and should perform well regardless of differences in demographics or the area of the state where they are trained and where they practice. If performance on an item is unrelated to general performance, or is interpreted and answered differently by subgroups of trainees, then that item is functioning poorly. Test items with poor performance were either modified and re-administered or removed on subsequent tests. *For an example of an analysis of test items, please refer to Appendix I for a copy of a Rasch analysis for a given test.*

[6] The database of multiple choice test questions is also referred to as an item bank.

Implementation Status
Implementation of the knowledge testing portion of the Framework has involved significant ongoing effort in developing and implementing evaluation tools and protocols.

Knowledge tests
Knowledge testing in three of the four standardized content areas (*Assessment, Case Planning, Placement and Permanency*) began in 2005 concurrently with the pilots of the Common Core. Knowledge testing in the fourth priority area (*Child and Youth Development*) began in 2007, concurrently with release of version 1.0 of this curriculum. After analysis of item performance and validity, CalSWEC developed and distributed multiple versions of modified tests for all content areas.

Item bank
CalSWEC continues to develop and maintain the item bank from which tests are constructed for the curricula that use knowledge tests. To date, approximately 275 items have been developed for the four content areas. Approximately 183 (or 66.6%) of these have been used in statewide tests. Of these, 153 have met standards for item quality and fit to current curriculum content and are active in the item bank. It is anticipated that an additional 20 multiple-choice test questions will undergo piloting (reflecting a total of about 74% of the item bank in circulation) before the end of FY 2009–2010. *(See Appendix J for a detailed report of the status of Item Piloting.)*

Infrastructure
In addition to purchasing item-banking and data-capture software, CalSWEC created scannable forms for each Common Core test and demographic form. With guidance from CDOG, CalSWEC also developed a protocol to collect feedback for development and refinement of test items from subject matter/content experts. *(See Appendix K for a copy of the Protocol for Revision of Problematic Knowledge Test Items for Common Core Training.)*

Reporting of results
CalSWEC reports results by region every six months, using a format developed and approved by CDOG. Some difficulties in both receipt and entry/analysis of the data were encountered earlier in the process, but these appear to be resolved and the system is operating successfully.

Resources Expended
Knowledge item development and analysis is a resource- and time-intensive process. At the local level, RTAs/IUC and counties expended staff time to review test items for the item bank and attend item-bank training. Staff time was also needed to prepare and distribute paper tests, and to manage storage and transmission of data. CalSWEC built infrastructure for data collection and analysis through the purchase of software and hardware to scan and generate tests. There were also costs associated with copying and mailing paper forms, and trainer time for learning test administration and transmission procedures.

Other resources expended by CalSWEC included staff time and associated costs for managing the process, item review, software training, and consultant services for item development, for item review and for software selection. Additional resources needed include assistance with statistical validation and scaling of items, data analysis and reporting, item bank maintenance, distribution of test templates, further item bank training and technical assistance, revising and/or writing new test items, and coordination of the data submissions.

Summary of Results

Overall learning

Data collected over the past four years have indicated that trainees improved significantly in their scores from pre-tests to post-tests. Improvements have been observed across regions and test versions for all content areas that include pre- and post-testing. The Common Core Curricula appears to build the knowledge of trainees, and the testing program implemented as part of the evaluation Framework demonstrates this.

For the *Critical Thinking in Child Welfare Assessment* curriculum, for which only a post-training test is administered, adequacy of performance must be assessed in relation to a pre-established standard of competence (i.e., a "cut off" score). At this time, no formal standard has been established that can serve as a yardstick across different trainee groups and test versions. However, analysis of the test data suggests that test items are performing well, and that trainees leave the classroom with significant knowledge related to the learning objectives: over 900 trainees have taken the most recent version of the test and the overall mean score is approximately 70% (a mean of 17 items correct out of the 24 analyzed).

Demographic factors

Analysis of the effects of demographic and other trainee-related variables on pre- and post-test score differences showed no significant differences in the degree of learning (i.e., the improvement from pre- to post-test) by the following characteristics: age, experience, educational level or type of degree, participation in the Title IV-E program, ESL, carrying a caseload, having heard that training was or was not worthwhile, having talked to a supervisor about training needs, or being able to think of specific cases to which training information could apply.

As with many standardized tests, some of the significant differences in performance that occurred were related to demographic characteristics. Specifically, there were greater differences in knowledge gained for Caucasian trainees in comparison to one or more other racial/ethnic groups in all three trainings, and greater differences in knowledge gains for women relative to men in two modules (*Family Engagement in Case Planning and Case Management*; *Child & Youth Development in a Child Welfare Context*). Such findings are not unusual; however, they merit further analysis and exploration. Significant pre/post differences also occurred related to RTA/region in the *Family Engagement in Case Planning and Case Management*, and *Placement and Permanency* trainings. Additional analyses are underway to identify any items that may be interpreted differently by trainees of different races or genders,

and a similar item-level analysis is being piloted in two regions to explore RTA-related differences. *(See Appendix L for a sample report of statewide knowledge test results.)*

Early analyses focused only on the effects of demographic variables on knowledge gains, controlling for initial pre-test level to parse the effects of differences coming in to training. However, the most recent analyses, conducted for the 2007 and 2008 calendar years, also estimated the effects of demographic variables on absolute level of performance at post-test. These analyses are generally of less interest in evaluating the effectiveness of training since those groups who have higher post-test scores also came in with more knowledge at pre-test, unless there is a significant effect of group membership on knowledge gained. One possible exception is the finding that trainees who report English as a second language scored significantly lower on post-tests. This merits further examination to ensure that test materials do not impose an undue language burden on these trainees.

Similar demographic findings were reported in an analysis by Los Angeles County researchers. In *Child & Youth Development,* "Education, race/ethnicity, and age emerged as significant predictors of performance"; in *Placement and Permanency*, "…education, race/ethnicity, age, and English as a second language were also significant predictors of performance" (Furman, Franke, Bagdasaryan, April 2009. "Specialized Preparation and Entry-Level Knowledge in PCW").

Possible Trainer Effects

An outgrowth of looking at trainee performance on test items is the possible effect of trainers on test performance. Prompted by concerns about lower-than-average test scores in one curriculum area, the Public Child Welfare Training Academy (PCWTA) undertook an examination of trainee performance on specific test items and competency areas. Specific knowledge test items/competency areas were identified where there were differences in test item difficulty depending on which trainer taught the material. The goal of this analysis was to provide more information to trainers and curriculum developers to help fine-tune course content and delivery.

Since the test items only sample from the whole content in a given area and don't cover every possible learning point, identification of these areas was intended to open discussion with trainers about areas where coverage of the course material could be improved and/or fidelity and consistency of delivery could be increased. Item score differences are a function of many variables including the trainee and the curriculum as well as the trainer. Therefore, these discussions were important in determining at what point addressing differences in training delivery might be important in improving trainee performance. The process was approved by the trainers, who all wanted such information for their own professional development. The academy conducted meetings to review findings where the trainers traded information about successful teaching strategies in key areas. The trainers also noted if and when they were not training specific content related to question areas. This analysis did not fully answer the initial question of why the academy was scoring lower than the statewide average, but pointed to areas in which training delivery could be improved to more closely follow the curriculum.

Title IV-E participation

The collection of demographic data has the benefit of allowing further analysis of how Title IV-E educational preparation relates to performance on the Common Core training tests. Although knowledge gains were similar for Title IV-E graduates and those with other types of trainees, Title IV-E graduates might be expected to achieve significantly higher scores at pre-test, and possibly also at post-test. This has largely been borne out. Statewide analyses for the most recent two years of data have shown a pattern of higher scores at both pre-test and post-test for Title IV-E-prepared trainees in the modules *Family Engagement in Case Planning and Case Management*, *Child and Youth Development in a Child Welfare Context*, and *Placement and Permanency*. Post-test score differences were statistically significant for two content areas: *Family Engagement in Case Planning and Case Management* and *Placement and Permanency*. Results for *Critical Thinking in Child Welfare Assessment: Safety, Risk & Protective Capacity* did not show an advantage for Title IV-E-prepared trainees.

Differences noted varied by year. The significantly higher post-test scores for *Family Engagement in Case Planning and Case Management* and *Placement and Permanency* occurred in 2008 data only. In 2007, differences were not significant. For *Child and Youth Development in a Child Welfare Context*, a significant difference in pre-test scores occurred in 2008, but the difference in post-test scores was not significant. In 2007, Title IV-E-prepared trainees scored higher at post-test, but virtually no differences occurred in scores at pre-test. Over time, these additional analyses will help identify any lasting patterns of differences in prior knowledge or training-related knowledge attained in specific competency areas related to whether or not trainees have received Title IV-E preparation. Any such differences in specific content knowledge may then be used to inform future curriculum development. *To see the complete analysis of CalSWEC Title IV-E and non-Title IV-E trainees, go to Appendix M.*

In Los Angeles County, Title IV-E participation "was not a significant predictor of performance" on the *Child & Youth Development* test (Furman, Franke, Bagdasaryan, April 2009). But with regard to the *Placement and Permanency* training, "significant differences by Title IV-E participation" were noted (Franke, Furman, and Bagdasaryan, April 2009). The authors speculate that the difference lies in the specialized content of the *Placement & Permanency* curriculum and the specialized training that is relevant to entry-level practice.

In a second analysis, Franke, Furman, and Bagdasaryan (January 2009) sought "to understand how variables simultaneously influence test performance…regression models were tested using demographic and experiential variables as predictors of test performance." In essence, many different demographic and educational variables related to the performance on different tests. The authors concluded, "Overall, the results from these models suggest that for many of the tests there are factors not measured that are likely more important predictors" of trainee test performance (Franke, Furman, and Badasaryan, January 2009).

Assessment of Progress

Knowledge test development, implementation, and analysis have expended the most resources and time in implementing the Framework. This makes sense, since assessing for knowledge is fairly straightforward and imparting knowledge is clearly one of the primary goals of the Common Core training program. Overall, the knowledge testing program has been successful. Although including content experts in development and refinement of items lengthened the process, it was time well spent—nearly all the items performed well and were validated on their first round of testing. Since the items performed well initially, they yielded sound information about whether the training was successful very early in the process.

Obviously, the training system must understand and address the significant differences in knowledge gains found for some racial/ethnic groups compared to others, as well as for gender and region. These findings warrant further analysis, determination of contributing factors, and implementation of appropriate remedies.

Future Directions

Ongoing evaluation of item performance will continue as part of the data analysis. Revisions of knowledge test items will continue as revisions to the curricula are completed. If curricula are stable over an extended period in a particular topic area, and results continue to show significant knowledge gains and fidelity of delivery, some knowledge tests may be suspended until revision of the curriculum. This could conserve time and energy to focus on other levels or aspects of evaluation.

Ongoing collection of linked knowledge test and demographic data offers significant opportunity for further research and evaluation. Since the identifier code is self-generated and stable over time, other instruments could be administered and compared to the data on knowledge acquisition. This potential for further research should be considered when making decisions about continuing or discontinuing tests, since data sets that extend over time often yield more powerful research findings.

Given resource constraints at this time, evaluation efforts at the knowledge level will not expand beyond the Common Core curricula already being evaluated. Instead, evaluations at the knowledge level will proceed in a more diagnostic direction in the next strategic plan. Since the previous strategic plan focused more on whether or not training appeared to be effective, transitioning to a diagnostic direction requires more targeted efforts. CDOG will be involved in weighting learning objectives for a given curriculum. This activity will refine the content focus of tests. Knowledge tests will be developed that reflect the weighting of the learning objectives, and the most heavily weighted learning objectives will have proportionally more items on the test.

For the foreseeable future, knowledge tests will continue for the following curricula: *Child and Youth Development in a Child Welfare Context*; *Critical Thinking in Child Welfare Assessment: Safety, Risk & Protective Capacity* (CAT county version); *Family Engagement in Case Planning and Case*

Management; and *Permanency and Placement*. Tests will be modified in conjunction with revisions to the respective curricula. Development and revision of test questions will occur as necessary to support more targeted evaluation efforts.

Analysis of differential functioning—where test items are analyzed with respect to demographic data to determine if bias appears to exist for a given test question—will continue. Race, gender, and ESL status will be explored via a pilot study with the Public Child Welfare Training Academy (PCWTA) to examine the possible effect of stereotype threat[7] in trainee test performance.

The next strategic plan will also explore trainer-level differences in item performance in order to provide feedback on fidelity of curriculum delivery. Performance differences for Title IV-E students compared to non-Title IV-E students will be monitored to better understand the variation that has been seen in relation to time period and Common Core module. As curricula are revised and stabilized, this analysis may yield more consistent patterns of results.

Level 5: Skills

Scope and Description

Embedded evaluations were specified in the Framework to provide information about the trainings' effectiveness in developing key casework skills. Use of embedded evaluations permits collection of data on specific skills in the classroom. Prior to completing the evaluation tools, trainees learn about the skill, see the skill demonstrated, and practice the skill. In the embedded evaluations in the line worker Common Core, trainees use standard instruments to assess whether child maltreatment had occurred or not in a fictive scenario. For a training module in the Common Core for Supervisors, trainees assess how a supervisor should respond to a fictive supervision scenario. Embedded evaluations enhance trainee learning and provide relevant feedback to trainers for course improvement, while also providing important data on trainees' acquisition of skills. This type of evaluation requires both standardized content and method of delivery.

These evaluations, like the knowledge evaluations described in Level 4, are used for course improvement and demonstrating competency of workers in the aggregate (not individually). CDOG chose to conduct these evaluations at the end of training only. Using performance tasks to test skills prior to training on a given topic is usually impractical since it is very time consuming and technically difficult and therefore costly. Moreover, few participants will enter training with the specific skill to be evaluated. CDOG also decided to evaluate competencies

[7] Definition of stereotype threat: "Is being at risk of confirming, as self-characteristic, a negative stereotype about one's group" (Steele & Aronson, 1995). This is a situational threat, meaning that it is not dependent on an internalized belief in the stereotype, just a fear that one will be judged by others through the lens of a given negative stereotype.

related solely to abuse (physical, sexual), and not neglect, for the embedded evaluations designed for new child welfare workers.

As in Level 4, RTAs/IUC and counties—which have access to statewide aggregate data and data from their own trainings—use results to determine the extent to which trainees are acquiring skills. As with the knowledge testing, the IUC provides information about trainee performance on the embedded evaluation to trainees' supervisors, so that supervisors may assist the trainees in their learning.

Implementation Status

Three content areas have been evaluated at the skill level. Embedded evaluations continue to be administered by the RTAs/IUC and analyzed by CalSWEC for the initial and revised versions of the Common Core Curricula with standard content *and* standard delivery for the following curricula:

- *Child Maltreatment Identification, Part I: Neglect, Emotional Abuse, & Physical Abuse* (CMI 1)
- *Child Maltreatment Identification, Part II: Sexual Abuse & Sexual Exploitation* (CMI 2)
- Common Core for Supervisors: *Casework Supervision* module

Additionally, an embedded evaluation was developed and is being finalized for the *Critical Thinking in Child Welfare Assessment: Safety, Risk & Protective Capacity* curriculum that can be used with the assessment tools used by Structured Decision-Making™ (SDM) counties in California. The embedded SDM evaluation represents a substitution for the previously developed knowledge test, which will be discontinued when the embedded evaluation is implemented. Expansion of the number of embedded tests has the same resource restraints noted above under knowledge tests. Pilots of this curriculum and evaluation tool are planned for FY 2009–2010.

An embedded evaluation was developed for the *Casework Supervision* module of the Common Core for Supervisors. The evaluation instrument and scoring guide have undergone several rounds of revision. The inter-rater reliability[8] on scoring was good for the most recent versions of this test.

Answer keys and scoring procedures were developed for all embedded evaluations, and a scoring rubric was developed for the *Casework Supervision* module of the Common Core for Supervisors. However, the Macro Evaluation Subcommittee has decided against pursuing cut-off scores at this time for tests that occur only at the end of training ("post-tests"), such as the embedded evaluations.

Resources Expended

Resources needed at this level included personnel time from the RTAs/IUC and the counties to participate in CDOG curriculum development activities, trainer/subject-matter-expert time to

[8] Interrater reliability is the extent to which two or more individuals (coders or raters) agree. Interrater reliability addresses the consistency of the implementation of a rating system.

consult on evaluation design and scoring rubrics, and trainer time to learn test administration and how to debrief the evaluations. CalSWEC staff and consultant time for evaluation design, analysis, and reporting to the RTAs/IUC, the counties, and the state were also needed.

Summary of Results

Both the CMI 1 and CMI 2 evaluations call for the trainees to judge whether or not each of a set of assessment factors, considered separately, would lead them to believe that abuse had occurred, then to make an overall judgment about whether or not abuse has occurred. Neither has formal cutoff scores for competent performance. However, in both cases a preliminary standard, established by CDOG, of 3 out of 4 correct on Part B (the judgment of whether or not abuse occurred) is used as a benchmark to describe trainee performance. As noted below, results consistently show that trainees generally leave the classroom able to demonstrate the critical skill of child maltreatment identification.

CMI 1

Results since 2005 have consistently shown strong performance on the overall decision-making component of the evaluation with approximately 90% of trainees statewide meeting the criterion of 3 out of 4 correct decisions. Trainees have been somewhat more likely to make incorrect decisions on non-abuse scenarios. The source of this difference is unclear and may be related to time allotted for administration at some training sessions, difficulty-level differences among the scenarios, or a general tendency of trainees to over-identify an issue rather than miss abuse that is present.

Analyses of the judgments on individual assessment factors in Part A have shown variability by factor and by scenario. This feedback has been used to modify the instructions for the evaluation, the wording of the evaluation items, the order of administration of the evaluation scenarios, and sections of the training curriculum to improve both the teaching of the content and the evaluation of course effectiveness. While some of these changes have been effective, error patterns overall have been very stable. Thus, efforts to improve performance on these factors are ongoing.

CM1 2

Since implementation of CMI 2 in 2007, overall performance has been strong on decision-making (whether or not sexual abuse occurred) with between 84% and 93% of trainees statewide passing using a criterion of 3 out of 4 correct decisions. Virtually no differences have occurred in percentage correct for abuse scenarios compared to non-abuse scenarios.

As with CMI 1, error patterns on Part A are being used to identify areas in which the evaluation materials need clarification and to inform curriculum revisions.

(See Appendix N for sample summary report of performance from CMI 1.)

Common Core for Supervisors: Casework Supervision Module
This embedded evaluation has undergone several revisions since its release in 2006. These revisions were made based in large part on trainees' difficulties with initial versions of the test. Subsequent revisions included clarifying a) the test scenario, b) the instructions, and c) the test questions. Additionally, over time it appeared that the emphasis of the content in this module was not delivered in a standard manner throughout the state. For these reasons, demographic analyses have yet to be completed; the "N" for a given test version simply was not large enough to determine whether or not statistically significant benefits or issues resulted from this training.

Assessment of Progress
Embedded evaluations require a significant amount of time in the training day. However, because embedded evaluations are both an evaluation and a training tool (via the debrief portion), they are a highly useful and effective means of assessing trainees in the training room. Embedded evaluations have provided valuable feedback for course revisions, as they highlight areas in which trainees are, or are not, grasping the concepts taught. In addition to highlighting areas of strength and weakness with respect to trainees' knowledge and skill acquisition, experiences with developing scoring keys and rubrics have highlighted areas in the curricula that require additional clarification and some areas that have inconsistent practice standards.

The experience of developing a scoring rubric for the Common Core for Supervisors embedded evaluation illustrates the importance of consistent practice and training standards. The specific skill evaluation focused on the teaching of an individual case conferencing model. Development of the scoring rubric initially highlighted differences in scoring of trainee performance in this particular skill. As more pilot data were analyzed, however, a deeper disagreement emerged in terms of what trainers emphasized in the training content based on the supervisory practice in their region. This is an example of an unintended consequence of training evaluation.

When planning an evaluation, one is often first forced to clarify and delineate the curriculum. Often the evaluation reveals a lack of specificity in the material itself, or a lack of standardization in its delivery. Clarifying the curriculum then engenders conversations about the desired practice, and whether regions, trainers, or administrators disagree about what is considered "best" or standard practice.

Future Directions
Embedded evaluations will continue for the four Common Core Curricula described earlier, until the Macro Evaluation Subcommittee determines a change in evaluation format. As with knowledge testing, the availability of a large database of demographics and other test data offers an important opportunity to advance an understanding of how workers and supervisors approach key job-related tasks.

Consistent with the next strategic plan, additional analyses are in progress to determine the effects of demographics and other trainee background differences on performance in CMI 1 and

CMI 2 that may offer promise for other embedded evaluations. The Public Child Welfare Training Academy (PCWTA) plans to pilot a neglect scenario as part of an embedded evaluation and study of the effects of demographics and other trainee background differences with respect to CMI 1 decisions (on whether or not physical abuse occurred in a given scenario).

This type of analysis also might be particularly useful to create a more detailed understanding of workers' decision making in relation to risk and safety. In collaboration with the University of North Carolina (UNC) School of Medicine, CalSWEC and several of the regions intend to link the CMI 2 data to the Child Forensic Attitude Scale (CFAS), a measure developed by UNC that assesses social workers' attitudes toward child sexual abuse disclosures. This project may bring valuable understanding to the impact of underlying values on identification of child sexual abuse.

The SDM™ version of the *Critical Thinking in Child Welfare Assessment: Safety, Risk & Protective Capacity* curriculum has recently undergone changes which will require piloting prior to statewide implementation. The embedded evaluation will have to be piloted simultaneously. . Both pilots are planned for FY 2009–2010.

The *Casework Supervision* module of the Common Core for Supervisors will be revised in the near future to better reflect current practice. Its embedded evaluation will also be revised to reflect the emphasis in content in this curriculum.

As with knowledge testing, analyses of item functioning in different regions or for classes taught by different trainers may provide information about regional practice variation and fidelity of training delivery. This information is potentially important in promoting high-quality practice throughout the state and reducing unplanned variations in the preparation new workers receive. Embedded evaluations of job-related skills also lay the groundwork for evaluating transfer of learning to practice by a) allowing both the estimation of how much of what is learned is retained and transferred and b) for providing a basis for understanding the effects of personal and contextual variables (e.g., motivation, supervisor support) on transfer.

Level 6: Transfer of Learning

Scope and Description

In order to promote best practices, all well-functioning training programs aim to impact transfer of learning (TOL), or transfer of skills, trained in the classroom to the work done by trainees in the field. Evaluation of transfer of learning offers important feedback about the extent to which practice change is taking place, the conditions that facilitate or impede change, and training's role in the change process. This—the essence of training—may be considered the basis of all training evaluation. Assessing TOL is very complex—since trainees perform the skills outside the training room, their personal characteristics and the organizational environment greatly impact the results.

When the Framework was initially conceptualized, none of the Common Core Curricula contained standardized skill development components that could serve to support the measurement of skill development and use of skills on the job. Activities in this area therefore focused both on curriculum and evaluation.
The Framework planned for two principal activities related to Level 6. The first involved the curriculum development process described in Level 2—development of suggested transfer of learning activities for select areas of the Common Core. The second involved design and implementation of an evaluation of the role of field training programs in transfer of learning. This evaluation occurred in two phases: Phase 1 focused on self-report of transfer and could be implemented in the short term; Phase 2 examined the relationship between skill development in the classroom and demonstration of that skill in a work product.

Implementation Status
The Central and Northern Regional Training Academies participated in Phase 1 of the evaluation project, which assessed the extent to which the provision of mentoring services:

- increased perceived transfer (by workers and their supervisors) of Common Core knowledge and skills;
- increased worker satisfaction with the job and feelings of efficacy; and
- contributed to improved relationships with supervisors.

The evaluation plan called for a quasi-experimental, comparison-group design. Newly hired social workers were divided into two groups: those that received field training services and those that did not. (The comparison group was possible in part because some counties participated in the field training program, and others did not.) Both groups rated their skills, the supervisory support they received, and their job comfort and satisfaction at the beginning and end of a six-month period. Their supervisors also rated the social workers' skills and the supervision they provided. If mentoring was effective, the evaluation should show a greater skill gain for the group receiving field training services than for the group not receiving the services.

The participating RTAs collected pre-test data from 141 caseworkers and 82 supervisors. Data collection for Phase 1 was discontinued after preliminary results were obtained in spring 2005. At that time, one of the participating field training programs underwent significant changes in focus, and this made continued collection of data unworkable under the original design. Significant problems also arose with collecting data from comparison sites due to staff turnover.

To deal with these issues, the evaluation team redesigned Phase 2 of the project into a retrospective case study of an intervention in one RTA, which focused on increasing the timeliness and quality of court reports. Workers and their supervisors attended training in court report writing, and supervisors and their managers participated in a range of group and individual field training sessions. The evaluation was conducted and completed as planned in fall 2006, with the exception of a pre- and post-intervention analysis of the actual court reports

produced by workers who received training and were supervised by supervisors who received the field training intervention. Although sample reports were collected and scored using a rubric developed for the project, too few were available to draw any conclusions. One barrier in completing this component was the movement of supervisors to different positions; only a few supervisors who had participated in the field training activities project continued supervising the worker throughout the process to the post assessment.

Resources Expended

Resource requirements at this level of evaluation included time spent by training coordinators and mentors at the participating RTAs 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 was 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.

Summary of Results

Phase 1

Preliminary analyses showed that participants in both field training models, which differed by RTA, were equally satisfied with their experiences. The following items received the highest ratings:

- "I trust my mentor (i.e., field trainer) to keep our discussions confidential."
- "I knew how to contact my mentor."
- "My mentor clearly explained his or her role to me."
- "My mentor is a positive role model for me."

Items dealing with coordination with the workers' supervisors (e.g., "My mentor has helped me strengthen my relationship with my supervisor"; "My mentor and supervisor work together to help me identify my training needs"; and "My mentor and supervisor work together to support my learning") received the lowest ratings.

Participants in both field training models rated their skills significantly higher overall at post-test than at pre-test. The amount of change in perceived skill varied with the skill rated. Which ratings changed most from pre- to post-test varied with the academy model, suggesting a relationship between what field-based trainers emphasize and which areas are impacted.

Four areas were rated by both academy groups as most difficult, even at post-test. These were "Effectively having clients complete case plans"; "Incorporating law and policy requirements"; "Implementing case plans"; and "Interviewing."

No significant changes occurred in ratings of supervision or job comfort. Participants who expressed greater intentions to continue working in child welfare were associated with higher ratings of supervisors, more confidence in their interpersonal skills, and higher satisfaction with the mentoring programs.

Phase 2
Focus group, interview, and survey data have indicated improvements in timeliness of court reports and increased use by supervisors of organizers such as a standard Court Control Log format in the unit and daily calendars for workers.

Feedback was mixed with respect to report quality. Supervisors felt that the emphasis on timeliness had resulted in more complete reports since workers were not attempting to fill them out at the last minute. Supervisors specifically mentioned changes in the Child Development section of the court reports: after attending the training in court reporting designed by the field-based trainers, they felt that workers were including more information in this section about developmental milestones. Some court officers reported seeing improvement in the Findings and Orders section of the report. In contrast, other court officers reported no improvements and an assistant county director commented that she would still like to see improvement in the writing and thinking processes.

Both workers and supervisors felt that the tools provided by the mentors had helped make the supervisors' review process more consistent and made it possible for one supervisor to pick up where another had left off when necessary.

Communications between the field trainers and supervisors and supervisors and workers improved. Communication with court officers remained more problematic; a lack of their involvement in early efforts to design the intervention was cited as a weakness and a lesson learned.

Assessment of Progress
The two evaluations completed at this level have provided a great deal of feedback about what makes field-based training successful. There are no active evaluations at this time for the transfer level; however, this remains a critical area for future evaluation efforts. Results of the field training program evaluations indicate that these approaches are very promising. Self-report data from surveys, interviews, and focus groups suggest that these programs are of value in increasing worker self-efficacy, and suggest that they can lead to behavioral changes in target areas. These evaluations also have resulted in a better understanding of the importance of the agency context in facilitating or hindering transfer of learning and TOL evaluation efforts.

While the information gleaned from these evaluations was valuable, more careful exploration of transfer of learning would strengthen the chain of evidence in the critical area of skill development. Specifically, the field would benefit from evaluations with more rigorous designs that document changes in on-the-job behavior by more objective measures, such as evaluation of work products like case plans, risk assessment instruments, and court reports.

These evaluations are complex. In addition to attempts to measure and control for possible explanations of worker behavior outside of the training arena, significant challenges exist in

identifying and collecting the correct information. Evaluations must focus not only on what is trained, but on what is supported by agency policy and practice. For example, the drop-down menus used in putting together case planning documents in the SACWIS system (CWS/CMS) do not require workers to use the principles for writing the types of quality case plan objectives taught in the Common Core. There is likely to be little incentive to demonstrate these behaviors unless the county or individual supervisor requires it, making an evaluation of transfer of learning in this area unlikely to show practice change. This underscores the need for more involvement from agency staff in the design of data collection strategies, and selection of practices and products that are closely aligned with agency policies and practices as the targets of transfer of learning evaluations.

Other Experiences in the State

TOL initiatives have not been limited to the Framework. For example, partner organizations within the IUC have developed TOL initiatives for new workers. CSU—Long Beach Child Welfare Training Centre and the Los Angeles County Training Section have developed the "Portfolio and Training Guide for New CSWs." Its purpose is to direct and support professional development of new Children's Social Workers (CSWs) during their first year within the Los Angeles County Department of Children & Family Services. The Portfolio is a reference document for CSWs, as well as for SCSWs and managers who supervise new workers, beginning with the CSW Core Training Academy and continuing through the end of the CSW's nine-month probationary period. The Portfolio is divided into three sections:

- *CSW Training Guide*
  - Core Practice Model Overview
  - Worker Self Assessment
  - Field Checklist
  - Enhanced Field Day activity outlines

- *SCSW Guide to Training for New CSWs*
  - Core Practice Competencies
  - CSW Core Training Academy Topics
  - Field Checklist
  - CSW Performance Work Plan - Performance Expectations
  - SCSW Case Conferencing Elements

- *Additional Documents and References*

Evaluation of the Portfolio is in its initial stages. The evaluation includes a continuum of CSW self-assessment and other evaluations, and SCSW surveys and evaluations conducted during the Core Academy and throughout the probationary period. Future evaluation will include analyses of trainee demographic data and training evaluation test data in relation to Portfolio CSW and SCSW assessments.

Future Directions
Evaluations of transfer of learning that assess actual work products, although challenging and resource-intensive, are essential in child welfare. Demonstrating practice change is a key piece of evidence for the effectiveness of training. At present, CalSWEC and its partners have existing skill-level evaluations that can provide a basis for planning future evaluations at the transfer level. These offer the potential for a clear link between what is taught in training and what is demonstrated on the job. Additional work is also required in identifying the barriers and facilitators to transfer associated with trainee characteristics, agency context, and training structure and design. The field is moving toward an expanded view of training as a continuum of staff development rather than solely a classroom activity; the need will continue for evaluation of expanded training models including coaching and mentoring. California stands in a strong position to contribute to the understanding of what makes an effective training and development model.

Initial evaluations in this area have pointed to the need for future evaluations of transfer of learning to focus on skills where a model of practice is in place supported by policy and widely accepted by practitioners. Priority areas identified in the PIP, or through the Peer Quality Case Reviews, or advanced training in a specific empirically supported practice might offer a starting point for the identification of a skill or skills on which to focus transfer of learning evaluation. It will be important to select an evaluation focus that can be linked clearly to skills taught and evaluated in Common Core training in order to evaluate the training's effectiveness. Without a baseline measure of learning, it is impossible to know how much of the practice observed on the job is related to participation in training and how much is related to individual differences in prior knowledge and experience, motivation, perceptions of supervisor support, or a number of other possible contributing factors.

As part of the next strategic plan, in FY 2011–2012 CalSWEC and regional partners plan to move forward with a feasibility study of transfer of learning evaluations as applied at a statewide level. This feasibility study will look at what is needed (data sources, data collection mechanisms, and methods of measurement) and logistics (timeframes, access, resources, and responsibilities). A summary of findings and recommendations will be provided to assist with decision making and next steps.

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. A first step in linking worker training to better outcomes for children and families is to build the supporting components necessary to demonstrate that training has an effect on worker practice. California has chosen to focus on

this necessary groundwork in the Framework to build a firm foundation for future efforts to evaluate at this level.

Implementation Status

The initial Framework focused on building capacity to work at this level by constructing rigorous measures of knowledge and skill outcomes of training. The ability to measure learning and establish a link between training and learning outcomes is a key prerequisite to measuring any effects training has on practice and—through changes in practice—on children and families. Before meaningful evaluation can take place at this level, much work remains to be done to link training to practice change and to link workers' training evaluation results with case-level information.

Resources Expended

As at other levels, resource requirements at this level of evaluation include staff and consultant time to plan and execute the evaluation. Significant resources already exist in the form of administrative data such as data on ASFA outcomes from the CWS/CMS database. It may also be possible to link to other initiatives such as the Peer Quality Service Reviews at the state level. However, taking advantage of the information available is a difficult task and requires significant resources. For example, evaluators would need to work closely with those familiar with the data systems to determine whether or not the desired outcome data are available and of sufficient completeness and accuracy and which variables capture it. Evaluation at this level would require that the system of self-generated identifiers be changed to allow a linkage between a worker's training evaluation data and case data. This would likely necessitate an additional review to assure protection of human subjects, as well as changes to evaluation protocols. Moreover, it is often difficult to identify which worker had responsibility for a case at the relevant time.

Future Directions

Much work still needs to be done to set up the chain of evidence that can evaluate at the level of outcomes. Even with the strong measurement of the links in the chain, factors still exist outside the control of the evaluators and the training system. It is insufficient to measure child- and family-level outcomes before and after training, and then conclude that training is responsible for any changes that occur. Too many other variables intervene (e.g., caseload size, personal motivation and ability, other initiatives occurring in the agency simultaneously, etc.) to allow any inferences to be made about the role of training. More sophisticated designs are necessary to establish links between the different levels of evaluation, and to control for as many intervening variables as possible. Such efforts would likely require engagement of research faculty from around the state to assist in planning and implementing the evaluation.

At present, continuing to evaluate at the knowledge, skill, and particularly the transfer level will strengthen the chain of evidence and contribute to the building blocks that allow eventual evaluation at this level.

As part of the next strategic plan, in FY 2010–2011, CalSWEC and its regional partners plan to move forward with a feasibility study to look at linking training to outcomes evaluation as applied statewide. This study will incorporate possible topics linked to research projects from the statewide research agenda in CWS. The feasibility study will also include the following factors: data sources, data collection mechanisms, and methods of measurement as well as timeframes, access, resources, and responsibilities. A summary of findings and recommendations will be provided to assist with decision-making and next steps.

IV. Analysis of Progress and Findings

Achievements:
When assessing the progress of California's statewide child welfare training evaluation efforts, one must consider the achievements in the context of the state's initial CFSR and PIP in 2002–2004. At the time the PIP was formulated, no standardized statewide training program existed; neither did a method to track trainee completion of various training programs. Although many entities provided high quality training across the state, the training was not evaluated systematically. No means existed to describe on a statewide basis what trainees were learning, how much they were learning, and whether or not the training they attended was sufficient preparation for taking on the daily responsibilities of child welfare social worker/child welfare supervision.

The development and implementation of the Framework allowed the state to assess the value of training with greater legitimacy and authority. All the Common Core curricula that were evaluated performed well, and appeared to enhance critical knowledge and skills for new child welfare staff. Evaluating training in accordance with the Framework also necessitated much more standardized statewide curricula, and enhanced the ability of the training system to strategically plan for differing levels of standardization for different parts of the Common Core.

Development and implementation of the Standardized Common Core Curricula and the Framework for Training Evaluation required the ongoing support of and feedback from stakeholders around the state. The process brought all levels of stakeholders closer to a unified understanding of the purposes and implications of training evaluations for training administrators, child welfare supervisors, trainers, curriculum developers, evaluators, and trainees. Collaboration with RTA/IUC and county partners with an eye toward the efficient use of statewide resources has been consistent. Stakeholder participation through the Macro Evaluation Subcommittee has offered valuable guidance for evaluation efforts and has strengthened evaluation designs and tools.

Statewide training evaluation activities have expanded considerably since the first discussions of a Framework: developing, revising, and finalizing item banks; piloting curricula and tests; and subsequent curriculum revisions. An item bank and test forms have been developed with the active participation of content area experts statewide and in accordance with relevant standards for test construction and validation published by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement and Education (AERA/APA/NCME, 2002). Protocols for test administration, test item review, and data transmission and reporting mechanisms have been developed and refined. Efforts to streamline statewide data collection, organization, and analysis continue based on ongoing feedback from training partners.

Challenges:
No statewide effort is without its challenges. Developing and conducting an effective training and evaluation program requires significant funding and the use of resources of time, people, and materials. Addressing the training and evaluation needs of multiple stakeholders and trainees has been and continues to be a balancing act. The final (and perhaps greatest) challenge remains: determining how best to link training and training evaluations to outcomes for children and families. Meeting this challenge will require significant investment in the various training evaluation levels as part of an overall chain of evidence.

California is now able to evaluate the overall success of its statewide training program and evaluation has become institutionalized as a part of training in the state. In the context of a complex and decentralized county-based system, the development of a sophisticated multi-level system of training evaluation remains a significant accomplishment. With a coordinated and collaborative system of evaluation, California is better able to meet the requirements of the CFSR. As a result, California is also in a better position to improve training content, delivery, and eventual implementation of practices, leading to improved outcomes for children and families served by CWS.

V. Future Plans for Evaluation

The Macro Evaluation Subcommittee convened via face-to-face meetings and conference calls in 2009 to finalize the next strategic plan for child welfare training evaluation in California, making several key decisions with respect to the plan. This section of the report describes the major decisions—both overarching and by level of training evaluation. *A more detailed description of future projects is outlined in a Matrix Summary (Appendix O) and Gantt Chart (Appendix P).*

Overarching Decisions

The Macro Evaluation Subcommittee made several decisions that cross all levels, including:

- Changed the timeframe for the new strategic plan to coincide with the PIP timeline.
- Expanded the purpose of the Framework beyond the summary function, moving in a more diagnostic direction.
- Evaluated feasibility of continued expansion of statewide evaluation of non-classroom training.
- Maintained the focus of the statewide evaluations on the Common Core Curricula.
- Explored the feasibility of expanding efforts to link training evaluation to program evaluation.

Decisions at Different Levels of Evaluation:

The table below summarizes decisions made by the Macro Evaluation Subcommittee that are specific to the levels of evaluation.

Level of Evaluation	Associated Project for 2009–2012 Strategic Plan for Child Welfare Training Evaluation
Level 1: Tracking Training—Tracking attendance/ demographics—Evaluate who attended.	These efforts will continue at the county level. Regional training academies will also participate in tracking trainee attendance. ▪ Common Core for Line workers: Demographic profiles and related analyses of line worker Common Core test data will continue. ▪ Common Core for Supervisors: Demographic profiles and related analyses of Common Core for Supervisors test data will commence. ▪ Analyses of Title IV-E trainee test data: Will also be completed as part of this next strategic plan. Such analyses include comparisons of Title IV-E status and time on the job and comparisons of Title IV-E training evaluation data with Title IV-E workforce/career path data. Additional questions to be answered include: In which content areas are Title IV-E trainees improving? In which content areas are Title IV-E trainees NOT improving? In which content areas do Title IV-E trainees come in with more knowledge?

Level 2: Formative evaluation of training courses (course level)—Evaluate what is intended to be taught in the content, how it is taught, and how it was developed and revised.	▪ Formative evaluations for observers (and separate ones for trainers) will be divided into assessments of content and assessments of delivery. ▪ Formative evaluation materials also will be developed for a new statewide venture: the e-learning platform.
Level 3: Trainee Satisfaction—Evaluate how satisfied the participants were with the training experience and their opinions about its usefulness.	These efforts will continue solely at the regional and county levels.
Level 4: Trainee knowledge acquisition—Evaluate knowledge gained as a result of training.	▪ Continue knowledge tests for the following curricula: *Child and Youth Development in a Child Welfare Context*; *Critical Thinking in Child Welfare Assessment: Safety, Risk & Protective Capacity* (CAT county version); *Family Engagement in Case Planning and Case Management*; and *Permanency and Placement*. ▪ Move in a more diagnostic direction: design more targeted knowledge tests, based on weighting of curriculum learning objectives and feedback from the Content Development Oversight Group. ▪ Develop and revise test questions as necessary to support more targeted evaluation efforts. ▪ Continue analysis of differential functioning (where test items are analyzed with respect to demographic data to determine if bias appears to exist for a given test question). ▪ Pilot study to look at the possible effect of stereotype threat in trainee test performance with respect to race, gender, and ESL-status (will be piloted at the Public Child Welfare Training Academy (PCWTA). ▪ Explore trainer-level differences in item performance to provide feedback on fidelity of curriculum delivery. ▪ Compare and monitor differences in performance for Title IV-E students vs. non- Title IV-E students to better understand the variation that has been seen in relation to time period and Common Core module.

Level 5: Skills acquisition by the trainee as demonstrated in the classroom—Evaluate skills gained as a result of training.	▪ Continue analyses to determine the effects of demographics and other trainee background differences on performance in CMI 1 and 2. ▪ Pilot the embedded evaluation of the SDM™ version of the *Critical Thinking in Child Welfare Assessment: Safety, Risk & Protective Capacity* curriculum (which includes the SDM™ Hotline tools). ▪ Convene a workgroup for and revise the embedded evaluation for the Casework Supervision module of the Common Core for Supervisors to better reflect current practice. ▪ Pilot a neglect scenario as part of an embedded evaluation and study the effects of demographics and other trainee background differences with respect to CMI 1 decisions re: whether or not physical abuse occurred in a given scenario *(to be piloted by the Public Child Welfare Training Academy (PCWTA).*
Level 6: Transfer of Learning (use of knowledge and skill on the job)—Evaluation is used to measure the extent to which a participant can transfer the learning from the classroom to the work setting.	▪ Conduct a feasibility study of transfer of learning evaluations as applied statewide. This feasibility study will look at what is needed (data sources, data collection mechanisms, and methods of measurement) and logistics (timeframes, access, resources, and responsibilities). A summary of findings and recommendations will be provided to assist with decision making and next steps.
Level 7: Agency/client outcomes—Evaluation is used to measure the degree to which training affects achievement of specific agency goals or client outcomes.	▪ Conduct a feasibility study of linking training to outcomes evaluation as applied at a statewide level. This feasibility study will incorporate possible topics linked to research projects from the statewide research agenda in CWS. The feasibility study also will include the following factors: data sources, data collection mechanisms, and methods of measurement; and timeframes, access, resources, and responsibilities. A summary of findings and recommendations will be provided to assist with decision making and next steps.
Other: *Attitudes/Values Evaluations* *Trainer Evaluation* *Quality Assurance*	▪ Attitudes/Values Evaluation re: CMI 1: The Public Child Welfare Training Academy (PCWTA) plans to pilot a neglect scenario as part of an embedded evaluation and study of the effects of demographics and other trainee background differences with respect to CMI 1 decisions (on whether or not maltreatment occurred in a given scenario). ▪ Attitudes/Values Evaluation re: CMI 2: In collaboration with the University of North Carolina (UNC) School of Medicine,

Marketing	CalSWEC and several of the regions hope to link the CMI 2 data to the Child Forensic Attitude Scale (CFAS), a measure developed by UNC that assesses social workers' attitudes toward child sexual abuse disclosures. ▪ Trainer Evaluation: Identify trainer-related differences in test item difficulty. Develop and obtain feedback on model of trainer evaluation. ▪ Quality Assurance: Convene small group of representatives from around the state. The same small group observes one Phase 1 training (curriculum up for revision) and one Phase 2 training (LOs up for revision) in each region. Analyze feedback from the small group of observers and provide information to Content Development Oversight Group and Statewide Training & Education Committee as part of curriculum/LO revisions and simultaneous trainer development activities. ▪ Marketing: Draft brief marketing piece(s) using non-technical evaluation language for targeted groups. Obtain feedback from Macro Evaluation Subcommittee and revise as necessary; disseminate first round of marketing piece(s) to stakeholders.

VI. Bibliography

Alliger, G.M., Tannenbaum, S.I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology,* 50, 341–358.

American Educational Research Association, American Psychological Association and National Council on Measurement in Education. (2002). *Standards for Educational and Psychological Testing.* American Educational Research Association: Washington, D.C.

Franke, T., Furman, W., Bagdasaryan, S. (January 2009). "Inter -University Consortium Core Academy Assessments, Fiscal Year 2007–2008". Annual Report.

Franke, T., Furman, W., Bagdasaryan, S. (April 2009). "Specialized Preparation and Entry Level Knowledge in PCW." Research Brief 3.

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, NY: Center for Development of Human Services.

Parry, C., & Berdie, J. (1999). *Training Evaluation in the Human Services.* Washington, D.C.: American Public Human Services Association.

Parry, C., Berdie, J., & Johnson, B. (2004). Strategic planning for child welfare training evaluation in California. In B. Johnson, V. Flores, & M. Henderson, (Eds.), *Proceedings of the Sixth Annual National Human Services Training Evaluation Symposium, 2003* (pp. 19–33). 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.

Steele, C.M. & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology, 69*(5), 797-811.

VII. Appendices

Appendix A: Levels of Training Evaluation

The AHA levels of evaluation are as follows:

Level 1, 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.

Level 2, ***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.

Level 5, ***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***, respectively, 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.

Appendix B: 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.

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

(Updated September 2008)

Level of Evaluation	Scope	Description	Decisions Pending	Resources	Status/ Timeframes
Level 1a: Tracking of Attendance	Completion of Common Core Training for line workers and supervisors will be tracked	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.	None	Databases (State, County, RTA) Personnel time to maintain and monitor (both locally and centrally) Protocols for individuals involved in maintaining and submitting data	Tracking will be in place July 2008 for Worker and Supervisor Common Core
Level 1b: Demo-graphics	All training	Demographic items were determined by the Macro Evaluation Team and a common demographics form is now developed and utilized for all training. Demographics are analyzed statewide and by RTA. Demographics are also used in analysis of two other levels of evaluation, e.g., knowledge and skills (embedded evaluations).	None	Databases (State, County, RTA) Personnel time to maintain and monitor (both locally and centrally) Protocols for individuals involved in maintaining and submitting data	Completed Data analysis is on going. Demographic analyses are currently conducted for statewide knowledge test data. They will be phased in for embedded skills evaluations as criteria for competent performance are finalized.

Level of Evaluation	Scope	Description	Decisions Pending	Resources	Status/ Timeframes
Level 2: Course (formative evaluation)	Phase 1/Big 7 of Worker Common Core and Supervisor Common Core	All courses evaluated by RTA teams using a Delta Plus instrument and CDOG+ discussion meetings. Revisions to curricula made based on formative evaluations.	None	Personnel time for: Reviewing evaluation results, ensuring quality and consistency e.g. observation/ monitoring delivery	Completed for the following versions: **Line Worker**: o Framework o Assessment o Child/Youth Dev o CMI Part 1 o CMI Part 2 o Case Planning/ Case Mgmt o Placement/ Permanency **Supervisor Core:** o Casework Supervision o Fiscal Essentials
Level 3: Satisfaction /Opinion	Phase 1/Big 7 of Worker Common Core and Supervision Common Core	RTAs and counties that deliver training 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 at individual RTAs/IUC and will continue.

Level of Evaluation	Scope	Description	Decisions Pending	Resources	Status/ Timeframes
Level 4: Knowledge	Phase 1/ Big 7	Knowledge item banks for four of the seven modules of Line Worker Common Core (excludes Framework which does not have a knowledge test and both CMI courses). Tests are constructed by CalSWEC from the item banks. Data analyzed on an ongoing basis. In 2008, IUC eliminated the knowledge test for the Assessment Module and piloted an embedded evaluation. Macro Evaluation Team will discuss IUC's pilot results and plan implementation of the Assessment embedded evaluation for California's SDM counties. Until a decision is made about whether or not counties using the original Critical Thinking Assessment curriculum want an embedded evaluation, these counties will continue to use knowledge tests in conjunction with the Assessment curriculum.	For counties using the original Critical Thinking Assessment curriculum (e.g., CAT counties), do they want to continue with knowledge testing or do they want an EE?	Staff time to prepare and distribute paper tests. Staff time for data management, input and transfer. Costs of scanner copying/ mailing Staff/consultant time for review and revision of knowledge items Staff/consultant time for analysis and reporting to RTAs, Counties, State.	Knowledge banks completed for current versions of modules being delivered now. Revisions of knowledge items will be made as revisions to Big 7 curricula are completed. Ongoing evaluation of item performance as part of data analysis.
Level 5: Skill during Training *Level 5: Skill during Training,*	1. Child Maltreatment Identification, Parts 1 and 2 2. Assessment Module	*CMI:* Embedded evaluations (EE) for these two modules focus on the ability to identify case information elements that do and do not suggest that abuse has occurred. The embedded evaluation for CMI 1 focuses on physical abuse and CMI 2 focuses on sexual abuse. Data are analyzed on an on-going basis. *Assessment:* Currently, IUC includes an embedded evaluation of skill. Will be moving towards implementation of embedded	CMI: Development of a minimum competency standard against which to judge performance. Assessment: Whether to develop an	Trainer/SME time for consulting on design and scoring rubrics Trainer time for learning administration and debriefing of evaluation Staff/consultant time for evaluation design, analysis and reporting to RTAs, Counties, State	CMI: Both evaluation formats are completed. Ongoing data analysis is being conducted. Assessment: IUC analyzing data. Implement across

Level of Evaluation	Scope	Description	Decisions Pending	Resources	Status/ Timeframes
cont'd	3. Supervisor Core	evaluation for SDM counties beyond IUC. *Casework Supervision Module:* Embedded evaluation on individual case conferencing model. Participants are given a written scenario summarizing what a hypothetical worker is telling them about a case. They respond to a series of questions that follow from the case conferencing model regarding what else they would like to know from the worker.	embedded eval for CAT counties. Supervision: Final content of scoring guide for roll out. Development of a minimum competency standard for performance.		SDM counties. Possible development of EE for CAT counties. Supervision: To date inter-rater reliability on scoring is good. One item on test has been revised and data from v1.3 will be reviewed again in Feb 09.
Level 6: Transfer	TBD	*Suggested transfer of learning activities for Case Planning Module*			
Level 7: Agency/ Client outcomes	TBD	*Using the entire framework, California can begin to build the 'Chain of Evidence' necessary to evaluate the impact of training on outcomes.*			

Appendix D: Standardized ID Code Assignment Instructions

California Common Core and Supervisor Core Curricula
Identification Code Assignment Instructions (Revised 7/1/06)

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 first three letters of your mother's *first* name? (Example: If your mother's maiden name was Alice Smith, the first three letters would be: **A L 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.**

   ___ ___ ___

3. 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 1st through the 9th, please put 0 in front of the numeral (example: **09**).

   ___ ___

   Combine these numbers to create your identification number (example: **SMIALI29**). 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.*

Appendix E: Demographic Survey (revised March 2009)

County Code: ☐☐ Trainee ID Code: ☐☐☐☐☐☐☐☐

RTA/IUC Code: ☐☐☐☐ Test Date: ☐☐ / ☐☐ / ☐☐☐☐
M M D D Y Y Y Y

42402

California Common Core Curricula Demographic Survey

Dear *Common Core Curricula* training participant,
By providing us 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 is kept entirely confidential. The information you provide us will not be associated with your identity or your performance in any way. Be sure to fill out the following survey once during the Common Core training. If you have already filled out this form, you do not need to do so again.

Choose the ONE BEST answer for each question and use an "X" mark to fill in the box corresponding to your choice below. Please do not put your answers (or any marks) anywhere else on this form. If you make a mistake, completely erase that choice and fill in the bubble next to the correct answer.

---

1) What is the HIGHEST level of your formal education? (Check only ONE box)

☐ High school ☐ MA/MS degree
☐ Some college ☐ MSW
☐ BA/BS degree ☐ PsyD
☐ BSW degree ☐ PhD - Field related to social work? ☐ Yes ☐ No

2) How long did you work in the field of child welfare prior to your current position?
☐ Under 6 months ☐ 6-11 months ☐ 1-2 years ☐ 3-5 years ☐ over 5 years

3) How long have you been in your current position?
☐ Under 6 months ☐ 6-11 months ☐ 1-2 years ☐ 3-5 years ☐ over 5 years

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 (if yes, go to 4a) ☐ NO (skip to question 5)

a) In which program did you participate?
☐ IV-E (LA DCFS) ☐ IV-E (CalSWEC) ☐ Other State

b) Were you in the child welfare field prior to your Title IV-E participation?
☐ YES (if yes, please answer 4c) ☐ NO (skip to question 5)

c) 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

a) If yes, which one?
☐ LCSW ☐ MFT ☐ Lic./Registered Psychologist ☐ Other: _____

Common Core Curricula Demographic Survey: v1.3 (Teleform Revision September 2009)
1

County Code: ☐☐ Trainee ID Code: ☐☐☐☐☐☐☐☐

RTA/IUC Code: ☐☐☐☐ Test Date: ☐☐ / ☐☐ / ☐☐☐☐
M M D D Y Y Y Y

42402

6) How do you identify yourself in terms of ethnicity/race? (Use an "X" for the appropriate space below.)

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

7) What is your age? ☐ 25 or younger ☐ 26-35 ☐ 36-45 ☐ 46 or older

8) What is your gender? ☐ Female ☐ Male ☐ Other (specify): _____

9) Is English your second language? ☐ Yes ☐ No

a) If yes, what is your first language: _____

10) Do you currently carry a caseload? ☐ Yes ☐ No

11) If yes, approximately how many children are on your caseload?
☐ 1-10 ☐ 11-20 ☐ 21-30 ☐ 31-40 ☐ 41-50

12) Which of the following statements best describes your feelings about attending Core?

- ☐ I am excited about attending the caseworker Core training and believe it will help me do my job better.
- ☐ I am unsure about what Core training has to offer me.
- ☐ The Core training is a requirement. I am looking forward to getting it over with.
- ☐ I have no feeling, either positive or negative, about attending the Core training.

13) I am concerned about the amount of time this class will require me to be away from my cases: ☐ Yes ☐ No

14) Prior to attending Core, I heard that this training was valuable/worthwhile. ☐ Yes ☐ No

15) Do you currently have a supervisor or mentor? ☐ Yes ☐ No

16) If yes, my supervisor (or mentor) and I have discussed my training needs. ☐ Yes ☐ No

17) I can think of specific child welfare children and families with whom I can use this training. ☐ Yes ☐ No

18) Please identify your professional role:

a. Are you a county Child Welfare Worker? ☐ Yes (If yes, stop here) ☐ No (If no, go to question 18b)

b. If you are not a county Child Welfare Worker, what organization do you serve?
☐ Tribal ☐ FFA ☐ Nonprofit ☐ Care provider ☐ Other (specify): _____

Common Core Curricula Demographic Survey: v1.3 (Teleform Revision September 2009)
2

Appendix F: Sample Delta Plus Tool (Course Level Evaluation)

SUPERVISOR CORE MODULE: FISCAL

- Delta Plus Change Evaluation Tool for Observers and Trainers of Supervisor Core Curricula -
(Starting December 2005)

Date of training: _____ **Name of observer or trainer:** _____ **RTA/IUC region or county:** _____

Day 1 of 1

INSTRUCTIONS: Please provide comments in the designated areas regarding any strengths or suggested changes that you feel are relevant to this training. This quality assurance/evaluation tool is designed to mimic the flow of the training day, in order to assist you in providing comments as the training day progresses. While you are not required to provide comments for every single segment of the training, we do appreciate any comments provided. **Please note if: 1) if not enough time has been allotted to cover a segment, and/or 2) a segment was not covered during the day specified on the lesson plan.** Your written feedback will assist us with future revisions of the California Supervisor Core Curricula.

Topic/Time	Learning Objective	Methodology	Strengths:	Suggested changes:
1:30 p.m. – 1:45 p.m. **Introduction to the Fiscal Section**	K3. The participant will understand various funding streams and fiscal implications. The three fiscal essentials that every supervisor should know are: • how the funding works for my program,	Welcome and introduction of trainer/participants, and review of session topics. Brief review of learning objectives. Lecture with PowerPoint. Introduce the three major areas to be covered: funding, allocations, and time studies.		

Topic/Time	Learning Objective	Methodology	Strengths:	Suggested changes:
	• how my county receives its share of the funds in the allocation, and how the time study process drives both the funding and the allocations.			
1:45 p.m. – **Federal Funding**	K3. The participant will understand various funding streams and fiscal implications.	Lecture with PowerPoint and notes from Trainer's guide: *-Title IVE* *-IVE Sharing Ratios* *-Discount Rate* *-Foster Care Admin Costs* *-CWS Training* *-Title IVB* *-Services under IVB* *-Title XIX* *-XIX Sharing Ratios* *-Title XX* *-TANF* *-TANF Sharing Ratios* *-Open ended Funding* Distribute the Social Security Act handout and have the participants review each of the titles listed. Have the participants complete the following exercise as a group. *Let us say that your county expenditures for Title IV-E case management total $100,000. How much would you be reimbursed*		

Topic/Time	Learning Objective	Methodology	Strengths:	Suggested changes:
		from the Feds? *Answer: $100,000 X 80% = $80,000 X 50% = $40,000* *Only 80% of the total foster care caseload is federally eligible based on our example and the feds reimburse 50% of eligible costs.*		
CWS Allocation	K3. The participant will understand various funding streams and fiscal implications.	Lecture from PowerPoint and notes in Trainer's Guide: *-Allocation History* *-CWS Budget & Allocation* *-Allocations* *-Allocation ltr CFL 05/06-25* Review the allocation letter hand out, CFL 05/06-25. Have the class take a few minutes to read the letter. Ask them to identify one or two premises. Go to attachment II and review justified FTEs, unit cost, etc. Go to attachment III and see the Federal revenue estimates. Go to attachment IV and review the augmentation amounts. Go to attachment V and review the program codes that earn your revenue.		
Time Studies	K3. The participant will	Lecture from PowerPoint and notes		

Topic/Time	Learning Objective	Methodology	Strengths:	Suggested changes:
	understand various funding streams and fiscal implications.	in Trainer's guide: -*Time is Money* -*Program Time Study* -*Supervisor Responsibility* -*Supervisor Importance* -*Program Code Descriptions* Review the handout CFL 05/06-26 Point out the changes for the December quarter. Next review the Social Services Function Program Code Description for 12/05 beginning with page 9. Review CWS-HR code 1441 thru 1444. Have the class read the description of what can be time studied to this code. Have them discuss what activities they think could be time studied here. Remind them that most of what they do is ensuring the health and mental health of children.		

Appendix G: Sample Global Evaluation Tool (Course Level Evaluation)

(next page)

RTA/IUC Code: ☐☐☐☐ Curriculum Title: _____

County Code: ☐☐ Observer Name: _____

Date of Training: ☐☐ ☐☐ ☐☐☐☐ Trainer(s): _____

M M D D Y Y Y Y

- California Social Work Education Center (CalSWEC) -

Global Evaluation Tool: Short Version for Trainers and Observers of California's Common Core & Supervisor Core Curricula

(Adapted from American Humane Association's Formative Evaluation Tool and Central RTA's and Bay Area Academy's Tools)

Instructions: Using a check mark ☑, please give each statement an overall rating of "Agree," "Somewhat Agree," "Somewhat Disagree," or "Disagree." If you check "Somewhat Disagree" or "Disagree," provide additional comments in the section below. (You can feel free to provide comments if you check off "Agree" or "Somewhat Agree," but commentary for these is not required. In addition, any comments on sections or segments that are well done are very much appreciated.) If additional space is needed, please feel free to use the reverse side of the page and/or additional pages.

A=Agree, SA=Somewhat Agree, SD*=Somewhat Disagree, D*=Disagree

A. CONTENT, PART 1 (LEARNING OBJECTIVES, LEVELS)	A	SA	SD*	D*
1. The competencies and learning objectives were clearly identified.				
2. The scope (breadth of coverage) and depth (detail) of content provide enough information to meet the learning objectives of the curriculum.				
3. The content is at the appropriate knowledge/skill level for the intended audience and is relevant to the job needs of the target audience.				

B. CONTENT, PART 2 (THEMES)	A	SA	SD*	D*
1. The content clearly communicated California policies and requirements.				
2. The training material addresses the child welfare outcomes of safety, permanence and well-being, and the impacts of these outcomes on CA data.				
3. The training integrated strengths-based practice concepts into the subject matter.				
4. The training integrated family and youth engagement practice concepts into the subject matter.				
5. The training integrated evidence-based practices and best practice concepts into the subject matter.				
6. The training integrated Fairness & Equity issues into the subject matter *(choose either "A," SA," SD," or "D" to provide ONLY an OVERALL rating of coverage of these issues)*: ❑ Disproportionate representation in CWS of African-American and Native American children (or children from other ethnic or racial groups, depending on county or region) ❑ Sensitivity to ethnic, racial, and cultural diversity; ❑ Socioeconomic status/poverty issues; Immigrant concerns; ❑ Sexual orientation (LGBTQ); ❑ Gender concerns; ❑ Disability issues; ❑ Age-related concerns; ❑ Rural area concerns; ❑ Other (please specify): _____				

RTA/IUC Code: ☐☐☐☐ Curriculum Title: _____

County Code: ☐☐ Observer Name: _____

Date of Training: ☐☐ ☐☐ ☐☐☐☐ Trainer(s): _____

M M D D Y Y Y Y

A=Agree, SA=Somewhat Agree, SD*=Somewhat Disagree, D*=Disagree

C. ORGANIZATION OF TRAINING	A	SA	SD*	D*
1. Related learning objectives were taught together and the relationship was clear.				
2. The training as a whole was introduced effectively.				
3. The content was sequenced from fundamental to more advanced.				
4. The most effective skills training utilizes the following sequence: description, modeling, practice, feedback. Was this sequence followed in this curriculum?				
5. The order in which topic areas were presented flowed well and made sense.				
6. Content areas where there may be differences in practices between counties were clearly identified.				
7. The time frames allowed by the curriculum were adequate for the training activities.				

D. TRAINING METHODS/DELIVERY	A	SA	SD*	D*
1. There was a sufficient mix of training methods to accommodate adult learning styles (e.g., visual, auditory, kinesthetic) and maintain interest.				
2. The training effectively incorporated stories, scenarios, or examples relevant to public child welfare to enhance content.				
3. Key concepts were woven throughout the entire curriculum and repeated in different contexts to reinforce retention.				
4. Clear instructions were provided for group activities.				

E. TRAINING MATERIALS	A	SA	SD*	D*
1. Trainee materials were easy to locate and follow.				
2. Participants were referred to materials in binder as a resource.				
3. Training materials reinforced training content.				
4. Training materials were clear, visually appealing, and varied in format.				
5. Training materials intended to be job reference tools were clearly identified & useful.				
6. Overheads/power point slides/tear sheets contained only key points or simple visual models.				
7. Audiovisual materials were appropriate in length, content and quality.				
8. Props (e.g., manipulatives, puzzles) were integrated into the content and designed for ease of use.				

F. EVALUATION ACTIVITIES	A	SA	SD*	D*
1. Adequate time was allowed for planned evaluation activities (pre- and post-tests, skills evaluation exercises, participant feedback forms).				
2. The purpose of the evaluation, use of ID codes, how the data will be reported and used, and confidentiality were clearly explained and participants' questions answered.				
3. Procedures for generating an ID code and completing the evaluation tools were clearly explained.				
4. Instructions to the trainer for introducing, conducting and debriefing (if applicable) of the evaluation exercises were clear and complete.				

RTA/IUC Code: ☐☐☐☐ Curriculum Title: _____

County Code: ☐☐ Observer Name: _____

Date of Training: ☐☐ ☐☐ ☐☐☐☐ Trainer(s): _____

M M D D Y Y Y Y

1. Areas of strength in this curriculum:

_____

_____

_____

_____

_____

_____

_____

_____

2. Areas for improvement in this curriculum:

1. *** = You MUST provide commentary if you checked either "Somewhat Disagree" or "Disagree."**
2. **Please provide an idea for improvement if you checked either "Somewhat Disagree" or "Disagree."**

_____

_____

_____

_____

_____

_____

_____

_____

3. Given concerns regarding time allotted for training, what, if at all, can be cut from this curriculum?

_____

_____

_____

_____

4. Additional comments *(can write more on the back of this page if necessary)***:**

Appendix H: Proposed Timelines for Revised Curricula/Evaluations (one curriculum at a time)

- DRAFT for 12/10/07 STEC Meeting -

(If the curricula were new, there would need to be more time up front to:
form the advisory committee, advisory committee to draft the competencies and learning objectives, finalize the competencies and learning objectives, and get them approved.)

April 2008 | May 2008 | June 2008 | July 2008 | August 2008 | September 2008 | October 2008 | November 2008

1. 3/20/08 – 4/1/08:
- Advisory committee forms (3/20/08).
- Advisory committee finalizes competencies, LOs, and content that the curriculum writer and evaluator are expected to address.
- Advisory team prioritizes edits based on feedback.

2. 4/1/08 – 6/30/08:
- Writers draft or revise curric.
- Evaluators and writers discuss eval design as part of curric. drafting process.
- Evaluators begin working on instruments.

3. 6/30/08:
- Writers send polished 1st draft of curric
- Evaluators send polished 1st draft of instruments

*to advisory team for feedback.

4. 7/1/08 - 7/31/08:
By 7/31/08, advisory team will have:
- Reviewed curric. & eval docs
- Communicated with each other & finalized feedback
- Sent summary of their comments back to writers/evaluators, to be included in final (1st pilot) draft.

5. 8/1/08 – 8/31/08:
- Writers and evaluators *(work together in cases where there is overlap)* incorporate and finalize respective edits to curric and eval docs.
- By 8/31/08, writers & evaluators must send final drafts of pilot curric. and eval docs to CalSWEC for final technical edits.

6. 9/1/08 – 9/30/08: Technical editor to make final edits to 1st pilot copy of curric. and eval docs.

7. 10/1/08: CalSWEC sends final 1st pilot copy of curric. and eval docs to:
- Printers
- Trainer of 1st pilot

8. 11/1/08: **1st Pilot of new/rev curriculum**
- Audience: SMEs, trainers, eval'ers
- Pilot respective evaluation
- Verbal feedback & summarize 1st pilot (formative & test) eval data on flip charts on site *(to be sent by 11/15 to advisory team)*

9. 11/15 – 12/15/08:
By 12/15/08, advisory team will have:
- Reviewed curric. & eval docs
- Communicated with each other & finalized feedback
- Sent summary of their comments back to writers/evaluators, to be included in final (2nd pilot) draft.

December 2008 | January 2009 | February 2009 | March 2009 | April 2009 | May 2009 | June 2009

10. 12/15/08 – 1/15/09:
- Writers and evaluators *(work together in cases where there is overlap)* incorporate and finalize respective edits to final versions of curric and eval docs.
- By 1/15/09, writers & evaluators must send final drafts of 2nd pilot curric. and eval docs to CalSWEC for final technical edits.

11. 1/15/09 – 2/15/09: Technical editor to make final edits to 2nd pilot copy of curric. and eval docs.

12. 2/15/09: CalSWEC sends final 2nd pilot copy of curric. and eval docs to:
- Printers
- Trainer of 2nd pilot

13. 3/1/09: **2nd Pilot of new/rev curriculum**
- Audience: **Actual CWWs**, plus SMEs, trainers, and evaluators
- Include respective evaluation
- Verbal feedback & summarize 2nd pilot (formative & test) evaluation data on flip charts on-site *(to be sent by 3/15/09 to advisory team)*

14. 3/15/09 - 4/15/09:
By 4/15/09, advisory team will have:
- Reviewed curriculum and eval instruments from 2nd pilot
- Reviewed evaluation data from 2nd pilot
- Communicated with each other & finalized team's feedback
- Sent summary of their comments back to writers and evaluators, to be included in final draft.

15. 4/15/09 – 5/15/09:
- Writers and evaluators *(work together in cases where there is overlap)* incorporate and finalize respective edits to final versions of curric. & eval docs.
- By 4/30/09, writers & evaluators MUST send final drafts curric. and eval docs to CalSWEC for final technical edits.

16. 5/15/09 – 6/15/09: Technical editor to make final edits to T4T/final copy of curric. and eval docs.

17. 6/15/09: CalSWEC sends final/ T4T copy of curriculum and eval docs to:
- Printers
- **Trainer of T4T**

18. 6/30/09: **Formal release and T4T** of new/rev. curriculum and evaluation docs for statewide trainers.

Appendix I: Sample Rasch Analysis of Test Items (Validation Process)

The tables and graph in this appendix illustrate the steps of the statistical item analysis using data from the most recent analysis of Placement and Permanency test data. Data included are from calendar years 2007 and 2008. Examples are provided of item difficulty and discrimination statistics as well as the process of identifying differential item functioning.

Item Difficulty and Discrimination: The following table shows the item analysis statistics for the Placement and Permanency post-test data. The item bank number is shown in the far right column. Items appear in order from the most difficult at the top of the table, down to the easiest. The column labeled "Score Corr." is the item discrimination statistic that shows the relationship of scores on each item to scores on the test as a whole. Negative or near zero correlations indicate that an item is not functioning well; either people who score high on the test as a whole are missing it more often that would be expected or people who score low on the test as a whole are getting it correct more often than expected. Negative discrimination values may mean that an item is confusing in some way; possibly having more than one correct answer or no clearly correct answer, or that there are errors occurring, particularly among items at the extremes of the difficulty range, due to guessing or carelessness. When items display negative or zero correlations with the test as a whole they are removed from the final analysis and flagged for further examination. In the table below item PL074 was excluded from the final analysis for poor discrimination. This item was also the most difficult of all items tested.

Note that although the individual test versions are not more than 30 items long, the table shows values for 54 items. Rasch allows items from different test versions to be put on a common scale, provided the versions have some items in common to provide a link. This in turn facilitates construction of a bank of items that may be combined in different ways to provide alternate forms of the test.

Item statistics

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	SCORE CORR.	ITEMS
45	82	474	2.89	.13	1.14	1.7	1.96	6.8	-.05	PL074
25	475	1982	2.46	.06	1.09	3.2	1.30	6.4	.15	PL071
11	489	1931	2.39	.06	1.08	2.7	1.28	6.3	.17	PL049
54	11	57	2.25	.34	1.03	.1	1.09	.4	.12	pL087
37	178	532	1.88	.10	1.04	1.2	1.12	2.2	.24	PL009
5	812	1988	1.58	.05	1.07	4.2	1.27	9.9	.22	PL030
41	224	532	1.47	.09	1.10	3.2	1.13	3.1	.19	PL023
30	668	1488	1.39	.06	.98	-.9	.99	-.5	.36	PL083
19	919	1988	1.33	.05	1.17	9.9	1.30	9.9	.12	PL039
15	685	1454	1.31	.06	1.06	3.2	1.13	4.9	.27	PL079
46	271	533	1.08	.09	1.10	3.6	1.15	4.0	.19	PL075
43	257	476	.98	.10	.96	-1.3	.95	-1.3	.38	PL026
8	1132	1982	.84	.05	.92	-4.9	.91	-4.5	.43	PL036
49	27	56	.82	.28	.96	-.6	.96	-.5	.32	pL081
47	275	474	.79	.10	1.02	.5	1.04	.9	.31	PL070

53	28	57	.77	.27	1.01	.2	1.02	.3	.23	pL080
3	896	1454	.64	.06	.94	-2.9	.92	-2.9	.41	PL089
38	298	476	.58	.10	1.01	.2	.99	-.1	.33	PL010
1	928	1455	.53	.06	.95	-2.3	.92	-2.6	.40	PL084
12	1301	1989	.44	.05	.99	-.7	.98	-.8	.35	PL018
42	332	476	.22	.11	.90	-2.1	.89	-1.9	.44	PL072
32	338	475	.15	.11	1.00	-.1	.95	-.8	.34	PL005
16	1077	1511	.13	.06	1.01	.2	1.03	.8	.32	PL076
44	339	474	.13	.11	1.11	2.1	1.03	.4	.21	PL013
28	1055	1437	.03	.06	.95	-1.7	.96	-.9	.38	PL027
10	1423	1930	.01	.05	.98	-.8	.98	-.5	.35	PL069
9	1435	1931	-.03	.06	1.00	.0	.96	-.9	.33	PL008
26	1128	1512	-.06	.06	1.06	1.9	1.03	.6	.26	PL040
40	404	533	-.16	.11	.91	-1.7	.90	-1.4	.41	PL015
23	1164	1512	-.20	.06	.91	-2.8	.85	-3.2	.43	PL022
14	1531	1982	-.22	.06	1.00	-.1	.97	-.7	.32	PL047
48	42	57	-.36	.31	1.03	.2	1.02	.1	.19	pL025
2	1586	1987	-.39	.06	1.16	4.4	1.23	4.4	.12	PL001
18	1563	1929	-.46	.06	.92	-2.2	.86	-2.8	.40	PL012
29	1212	1492	-.49	.07	.99	-.3	.98	-.4	.32	PL073
24	1616	1988	-.49	.06	.97	-1.0	.99	-.1	.32	PL028
31	389	475	-.52	.12	.93	-.9	.95	-.5	.36	PL021
17	1252	1511	-.62	.07	1.07	1.6	1.05	.8	.22	PL078
34	439	531	-.62	.12	.90	-1.4	.92	-.8	.38	PL029
6	1262	1510	-.67	.07	.91	-1.9	.84	-2.5	.39	PL077
21	1219	1454	-.68	.08	1.08	1.7	1.08	1.1	.21	PL085
22	1325	1511	-1.04	.08	.88	-2.1	.78	-2.8	.40	PL090
36	419	476	-1.05	.15	.82	-1.9	.72	-2.2	.48	PL041
35	473	532	-1.18	.14	.83	-1.8	.60	-3.2	.48	PL038
33	428	474	-1.30	.16	.80	-1.8	.53	-3.4	.51	PL007
50	51	57	-1.52	.44	.96	-.1	.78	-.6	.30	pL082
7	1340	1457	-1.54	.10	.87	-1.7	.70	-3.0	.38	PL086
13	1400	1513	-1.64	.10	1.01	.2	.94	-.5	.23	PL046
27	1781	1912	-1.72	.09	.85	-2.1	.56	-4.9	.42	PL045
20	1362	1456	-1.80	.11	.84	-1.9	.62	-3.4	.40	PL033
39	450	476	-1.96	.21	1.06	.3	1.37	1.4	.08	PL035
52	53	57	-1.98	.52	1.01	.0	1.05	.1	.13	pL034
4	1382	1453	-2.10	.13	.84	-1.6	.65	-2.6	.37	PL011
51	53	56	-2.28	.60	1.01	.0	1.19	.3	.07	pL024
MEAN	764.	1102.	.00	.13	.98	.0	.99	.1		
S.D.	535.	694.	1.25	.12	.09	2.4	.23	3.2		

Differential Item Functioning: Differential item functioning is of concern when trainees' probability of getting an item correct varies significantly depending on trainee characteristics that should be irrelevant to success. Thus, differences in item difficulty for different racial or ethnic groups, trainees from different areas of the state, or for different genders would be of concern. Differences in success rates due to differences in such things as education or experience would be expected, and would not be of concern. In fact, these differences could be taken as evidence of scale validity.

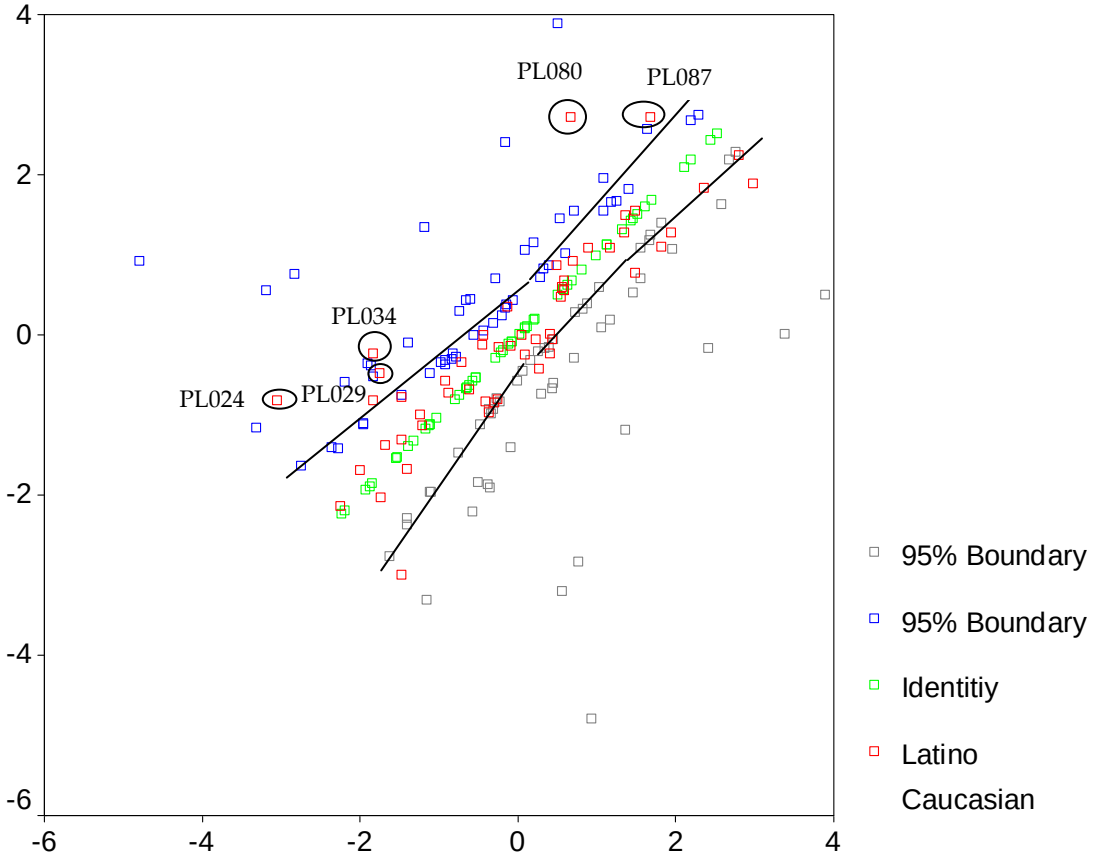
The process for determining whether or not difficulty of an item varies involves deriving separate sets of item difficulty estimates for the groups being compared. The degree of overall similarity in item

difficulty is examined through correlating two sets of item difficulties. High correlations indicate that items that are difficult for one group of trainees are also difficult the other group and vice versa. Correlations of .80 and above indicate excellent agreement.

Even with good overall agreement, differences can occur for individual items. Individual items that differ significantly in difficulty for a focal group (e.g. males) are identified by graphing the two sets of difficulties against each other. The scatterplot below shows an example of this type of graph using placement and permanency post-test data for Latino trainees and Caucasian trainees .

Each item is represented by a red square where the difficulty values derived for the Latino trainees and Caucasian trainees' values intersect. Items with similar difficulties fall on, or close to, the green diagonal line. The curved lines formed by the blue and gray squares indicate a confidence interval of 95%. Any points that lie outside these lines represent items where the difficulty based on one group is significantly different from the difficulty based on the other (i.e. There is a less than 5 in 100 chance that the difference in difficulty is just due to chance).

For the data shown in the example, items PL087, PL080, PL034, PL029 and PL024 are significantly more difficult for Latino trainees than Caucasian trainees. The overall correlation for the two sets of difficulty estimates is .86.



The following items were excluded from the analysis of trainee performance based on concerns with item functioning. The item numbers, reasons for exclusion, and recommendations for use in future test versions are shown in the table below.

Item Excluded	N of Responses	Test Version(s)	Reason	Recommendation
PL074	474	1.0	Negative Discrimination	Examine for need to modify or drop
PL010	476	1.0	Possible DIF[9] based on race	Examine for need to modify or drop
PL013	474	1.0	Possible DIF based on race, gender	Examine for need to modify or drop
PL023	532	1.0, 1.15	Possible DIF based on race	Examine for need to modify or drop
PL024	56	1.15	Possible DIF based on race	Defer decision. Easy item, relatively few responses.
PL034	57	1.15	Possible DIF based on race	Defer decision. Easy item, relatively few responses.
PL075	476	1.0	Possible DIF based on race	Examine for need to modify or drop
PL080	57	1.15	Possible DIF based on race	Defer decision. Relatively few responses.
PL087	57	1.15	Possible DIF based on race, ESL status	Examine for possible vocabulary, readability issues
PL029	531	1.0, 1.15	Possible DIF based on race	Examine for need to modify or drop

[9] DIF is Differential Item Functioning. It refers to a situation where a disproportionate number of trainees from a focal group (e.g. based on race or gender) miss the item in comparison to other trainees of similar overall ability.

Appendix J: Item Piloting Update (March 2009)

Module	Test Versions	Number of new items tested	Test characteristics	Evidence of Learning	Number of items that meet performance criteria/current curriculum fit	Items retired or inactive due to performance or content issues	Items in bank remaining to be tested
Assessment	1.0	30	Rel. .65		50	7 Retired/inactive 2 Proposed for further review based on content	8
	1.1	18	Rel. .52				
	1.15	1	Not analyzed separately				
	1.16	1	Rel. 53	Average correct 71%			
Case Planning	1.0	25	Rel. =.65	Statistically significant pre-to post-test improvement	40	3 Retired/inactive	17
	1.2	16	Rel. =.57				
	1.25	2	Rel. =.63				
	1.26	1	Rel. = .60				
Child and Youth Development	1.0	30	Not analyzed separately	Statistically significant pre-to post-test improvement	21	18 Retired/inactive	65
	1.05	5	Rel. =.56				
	1.1	20	No info yet		20 (to be tested)		39 of 80 have not been used
Placement and Permanency	1.0	30	Rel. =.56-.62	Statistically significant pre-to post-test improvement	41	0 Retired or Inactive 12 Proposed for further review based on performance	9
	1.1	17	Rel. = .56 -.60				
	1.15	7			1		1

Appendix K: Protocol for Revision of Problematic Knowledge Test Items

CALIFORNIA SOCIAL WORK EDUCATION CENTER
UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF SOCIAL WELFARE

Protocol for Revision of Problematic Knowledge Test Items for Common Core Training *(FINAL, 2-15-08)*

ISSUE:
In July 2007, the Macro Evaluation Team developed a protocol for addressing items that the RTAs/IUC had concerns about **multiple choice test items for Common Core curricula**. This protocol involved the RTA/IUC identifying the concerns to CalSWEC (via Leslie Zeitler) and then a process of looking at information generated by statistical analyses of item performance, followed by a CAT/CDOG+ group decision to keep testing the item (i.e., include it in test versions) or to remove it from testing (and either drop it or rewrite it – the latter going into a holding bin, or "purgatory"). We have found that this process may take too long and so the following protocol is being proposed for discussion and a decision by the Macro Evaluation Team.

BACKGROUND:
One of the goals for each curriculum is to have a high quality knowledge test, i.e., a valid and reliable set of multiple-choice test items from which tests can be constructed. There is an ongoing, multi-year process to reach this goal.

Currently each curriculum has a bank of multiple choice test items (at least five items per knowledge learning objective). These items were developed by CalSWEC, consultants, and respective CAT Teams (*now known as "CDOG+ group" for a given curriculum*) based on the curricula. Each multiple choice test item has been reviewed by the relevant CAT/CDOG+ group to make sure it is covered in the curriculum, that it is correct, and that it is clearly written. Each test version administered by the RTAs/IUC is made up of a subset of these multiple choice test items.

Once test results are in, CalSWEC and its consultants statistically analyze how these items are performing. This analysis is one means by which we learn whether each of the items is performing well AND (for most curricula) whether trainees are gaining knowledge as a result of the training. Over time (and given a large enough number of test takers), this statistical analysis helps identify which items should be kept and which need to be rewritten or dropped from the active item bank.

Even with the CAT/CDOG+ group review of items as they are being developed, actual delivery of training and administration of the test may lead to concerns about some of the items. For example, an item might have incorrect information or more than one answer may be equally correct. It is important that RTAs/IUC spot these items and that a decision is made about what to do with them in a timely way. This is the second way of getting feedback about items.

Given that RTAs/IUC staff sometimes identify problems with test items for curricula and that the concerns must be addressed quickly, the following protocol is proposed.

Protocol for Revision of Problematic Knowledge Test Items (FINAL, 2-15-08)

STEP 1:
RTA/IUC staff identifies problem items and sends an email to CalSWEC (Leslie Zeitler) describing why the item is of concern and any detailed suggestions about how to fix the item.

STEP 2:
CalSWEC reviews the RTA/IUC concerns using the following criteria:

Criterion A: The item has incorrect information either in the stem (the question) or in the answers (e.g., a subject matter expert says that the information in the item is incorrect, even if it reflects the curriculum content).

Criterion B: The necessary background information for the item isn't covered in the curriculum or is insufficiently covered and thus the trainee shouldn't be expected to know the answer or reason towards the correct answer.

Criterion C: More than one answer could be equally correct.

Criterion D: The wording of the item is unclear such that it is difficult to understand what is being asked.

It should be noted that there are two other possible reasons that might be given for concern and that these must be handled in a different way because, while important, they are not sufficient for making a decision in the same way that criteria A-D are.

Criterion E: Many test takers are not answering a given question correctly. It is important to remember that this test is not intended to measure whether a prescribed level of mastery is attained by a certain percentage of people; rather it is intended to measure whether learning is occurring, i.e., are test scores improving (for those where a pre-and post-test is given). These are different paradigms. So, the fact that, for example, only 50% of test takers are getting an item correct on the post-test is not in and of itself a reason to have concern about an item.

These items will not go immediately to "purgatory", if no other concerns with accuracy, clarity, or relationship to curriculum are identified, as it is necessary to have some more difficult items to avoid ceiling effects that limit the ability to show improvement. Their performance will be tracked as additional data becomes available.

Criterion F: The background for the material is in the curriculum, but the material isn't being covered in a particular training. In this case, it is important to explore whether the material

should be covered in training delivery and isn't, or alternatively, whether the question reflects training material that is justifiably not being covered in training, e.g., because there is too much material. This is a CAT/CDOG+ group issue regarding curriculum first and only then the item.

In Step 2, CalSWEC will review the RTA/IUC's concern. In cases where CalSWEC agrees that the item should be removed from testing (and either dropped or reworked), CalSWEC will let the RTA/IUC know and will inform the CAT/CDOG+ group by email, explaining the reasons and asking CAT/CDOG+ group members to respond with agreement (or not) by a certain deadline date. If by that deadline date there are no CAT/CDOG+ group members who are in disagreement with the decision to remove the item(s), CalSWEC will remove the items from the active test bank and put in them in purgatory for future rewrite (or will drop the item if it appears not to be fixable). CalSWEC will adjust the active item bank and ensure that test versions do not include these items.

If the issue with the item(s) of concern cannot be resolved via email:
If one or more CAT/CDOG+ group members does not agree with the decision to remove an item from the active test bank, then CalSWEC will arrange a CAT/CDOG+ group conference call to make a final decision. And, if CalSWEC does not agree with the concern expressed about a given item(s), CalSWEC will arrange for a CAT/CDOG+ group conference call.

STEP 3:
For items that are removed from the active test bank and are put in purgatory, CalSWEC and consultants will look at statistical data about how the item is performing and give that to the CAT/CDOG+ group before the item is rewritten. Also, if needed, other subject matter experts will be consulted.

STEP 4:
CAT/CDOG+ groups and CalSWEC consultants will periodically work on rewriting items. Item writing will undergo the same CAT/CDOG+ group scrutiny as has been done in the past. The items will be based on the curriculum and the learning objectives and the CAT/CDOG+ group will review the items until CAT/CDOG+ group approval is achieved. Only then will the items be added to the active item bank.

Note about Curriculum Stability and Test Items*: In addition to the above protocol, whenever curricula are undergoing rewrites (such as is the case now with Child and Youth Development), all items, whether identified as problematic or not, will be reviewed once the curriculum is finalized.*

Appendix L: Summary of Knowledge Test Performance (one curriculum)

CALIFORNIA SOCIAL WORK EDUCATION CENTER
UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF SOCIAL WELFARE

Statewide Report: Case Planning Knowledge Test
- March 2009 -

ITEM ANALYSIS

Data:

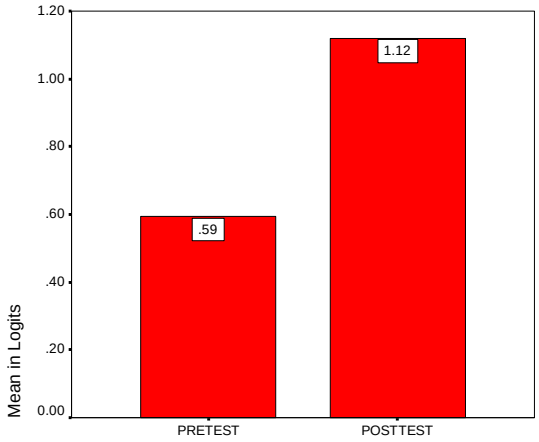
- 1,952 Responses
  - Version 1.0 N= 390
  - Version 1.25 N= 669
  - Version 1.26 N= 893
  - 407 pre-and post-tests could not be matched.
- Date range from January 2007 through December 2008.

Item Functioning:

- All except one item have been carried over from version 1.25. Item CP006 has been rewritten to be more precise and accurate and is being piloted for the first time as item CP006a.
- Item CP007 (which appears on v1.0) was excluded from the analyses based on early reported problems with its content.
- All other items met criteria for appropriate difficulty and discrimination (ability to differentiate among higher and lower ability test takers)
- Comparisons over two years of data did not show differences in difficulty between scenario and non-scenario items.

TEST RESULTS

Pre-test–Post-test Differences Calendar Year 2007



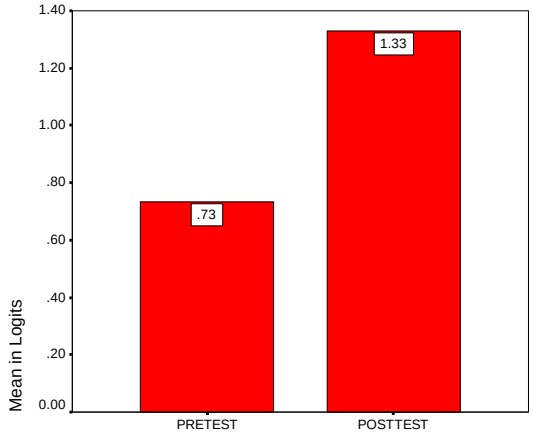
2007 Test Data

- 800 pairs of pre and posttests
- Data from 5 Academies/IUC
  - BAA N= 95pairs
  - Central N= 101 pairs
  - IUC N= 148 pairs
  - Northern N= 131 pairs
  - Southern N= 326 pairs
- Test versions 1.0, 1.25, 1.26

2007 Results

- Average gain of .53 logits
- Raw score change from 13.8 to 15.3 out of 24 points for version 1.0
- Raw score change from 15.5 to 18.2 out of 25 points for version 1.25
- Raw score change from 15.7 to 18.0 out of 25 points for version 1.26
- Statistically significant t=20.610 d.f.=801, p<.001

Pre-test–Post-test Differences Calendar Year 2008



2008 Test Data

- 743 pairs of pre and posttests
- Data from 5 Academies/IUC
  - BAA N= 119 pairs
  - Central N= 111 pairs
  - IUC N= 146 pairs
  - Northern N= 139 pairs
  - Southern N= 228 pairs
- Test versions 1.25, 1.26

2008 Results

- Average gain of .60 logits
- Raw score change from 16.1 to 18.7 out of 25 points for version 1.25
- Raw score change from 15.6 to 17.6 out of 25 points for version 1.26
- Statistically significant t=22.070 d.f.=742, p<.001

Demographics for 2007 and 2008[10]:

- Demographic forms were available for 1121 of the 1546[11] pairs of pre- and post-test tests overall (72.5%).
- Since many of the individual demographic variables are interrelated (e.g. carrying a caseload and concern for time away from the office, or educational level and Title IV-E status), consideration of the relationship of a single demographic variable to test scores can be misleading. To minimize this potential, a multiple regression analysis was done to assess the effects of a demographic variable on post-test score over and above the effects of other related variables.
- Pre-test is included as a predictor in this model to control for level of knowledge prior to the training. This allows an assessment of the relationship of background variables to learning (and training effectiveness) rather than to the overall level of knowledge shown on the post-test.
- Amount learned did not differ significantly by age[12], experience, educational level or type of degree, participation in the Title IV-E program, ESL, carrying a caseload, or expectations of Core.
- Several statistically significant differences were also noted in this analysis :
  - ⇨ Amount of positive change from pre-test to post-test and overall post-test level varied significantly by race. Hispanic /Latino trainees changed significantly less than Caucasian trainees ($p<.01$) and scored lower overall at post-test ($p<.001$). This difference remained after controlling for ESL status in the analysis. African American, Asian, and trainees in a combined group consisting of "American Indian/Alaska Native", "Multi-racial", and "Other" also scored significantly lower than Caucasian trainees at post-test ($p<.001$, $p<.05$, and $p<.01$, respectively), although the amount of pre-to post-test change was not significantly different for either group from the amount experienced by Caucasian trainees.
  - ⇨ ESL status was not significantly associated with amount of learning from pre- to post-test; both ESL and non-ESL participants gained knowledge in approximately equal amounts. However, ESL was significantly associated with lower overall post-test scores ($p<.001$).
  - ⇨ Women learned significantly more from pre- to post-test than men ($p<.01$). Female trainees also scored significantly higher on post-test than male trainees ($p<.001$).
  - ⇨ Participants from one of the Academies changed significantly less from pre-test to post-test than the others ($p<.01$) and also scored significantly lower at post-test ($p<.01$). Participants from another academy changed significantly more from pre-test to post-test than the others ($p<.05$) and also scored significantly higher at post-test ($p<.01$).
  - ⇨ Trainees who had MSWs and participated in Title IV-E programs scored significantly higher at post-test ($p<.001$). However, this group also experienced positive change from pre- to post-test that did not differ significantly from other trainees'.

Next Steps:

- Pilot all items identified by the CDOG+ Team as relevant.
- Enter well functioning items that are identified by the CDOG+ Team as relevant to the revised curriculum into the Examiner item bank.
- Work to increase the percentage of participants whose pre- and post- scores can be matched and whose test scores can be matched with their demographic data.
- Conduct further item analyses to identify areas of differential functioning with regard to race/ethnicity, gender, region, and ESL.

---

[10] The timeframe for the demographic analysis has been expanded for this report in order to capture all backlogged data.
[11] One test pair was not included in the calendar year analyses due to a missing date. Thus, the number of pairs is 1,546 rather than 1,545.
[12] Variables with multiple categories were recoded as dummy/contrast variables.

Appendix M: Summary of Test Performance by Title IV-E/non-Title IV-E Status

CALIFORNIA SOCIAL WORK EDUCATION CENTER
UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF SOCIAL WELFARE

Report to the CalSWEC Board: Analysis of Common Core Data, May 2009

BACKGROUND:
In 2002, CalSWEC and the California Regional Training Academies (RTAs)/Inter- University Consortium (IUC) began development of a statewide evaluation of Common Core training. This evaluation is part of the strategic plan for multi-level evaluation of child welfare training in California. 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. In doing so the strategic plan is directly responsive to the California Department of Social Services (CDSS) Program Improvement Plan (PIP).[13]

The development of the Common Core evaluations is following a rigorous process to ensure that test items reflect the competencies, learning objectives and content of the curricula and measure trainee learning as accurately and consistently as possible. Knowledge test plans, test items, and embedded skills assessment tools have been developed and critiqued by teams representing the RTAs/IUC, CalSWEC, counties, and consultants. Test items and supporting materials have undergone multiple revisions by these teams prior to use in testing situations. Some will undergo more revision as the teams review test findings. Results reported here are based only on items that have met review and performance standards; however, it is important to remember that both the curricula and tests are in a pilot stage and results may vary slightly from year to year or as different versions of tests and curricula are implemented.

RESULTS FOR 2007 AND 2008[14]:
Results are shown here for pre- and post knowledge testing in the Case Planning, Child and Youth Development, and Placement and Permanency modules, and for the embedded skills assessment in Child Maltreatment Identification, Parts I and II. Trainee profiles are also included which show select demographic, educational, and experience differences for trainees during this time period.

Overall, trainees are learning information identified as important by expert teams representing the RTAs/IUC, CalSWEC, and counties. Their knowledge from pre- to post-test increases at a statistically significant level.

[13] "Training Evaluation Report", California Social Work Education Center, December 20, 2004

[14] The timeframe for the analyses has been expanded for this report in order to capture all backlogged data. Data included is from tests administered between 1/1/07 and 12/31/08.

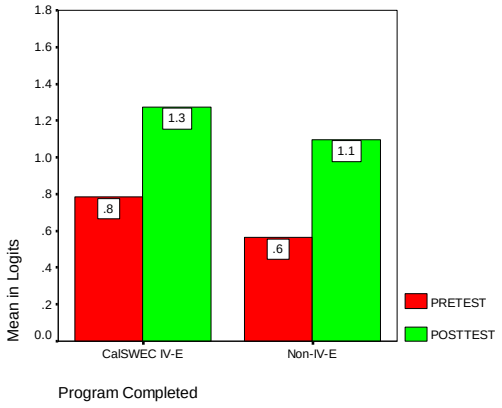
HOW ARE THE EVALUATION FINDINGS BEING USED?

- Knowledge test results are being used to show whether or not participants are learning important facts, concepts, procedures and relationships.
- Information on how trainees respond to individual items is being used to identify areas in the curricula that are unclear or need to be expanded or updated, as well as to identify test items that are too easy, too difficult, or confusing.
- Demographic and other background variables are used both to provide profiles of the workforce participating in Core training, and to aid in identifying issues with individual test items. Differences in the percentage of people from different groups who answer a test item incorrectly may reflect important differences in local practice, the way curriculum is being delivered regionally, or how trainees of different races or genders understand a question's content. These differences are tracked as part of the test validation process and items that are problematic will be rewritten or dropped.

- RESULTS FOR SELECT COMMON CORE TOPICS -

*CURRICULUM TOPIC: FAMILY ENGAGEMENT IN CASE PLANNING AND CASE MANAGEMENT

Pre-test–Post-test Differences by Title IV-E Participation Calendar Year 2007



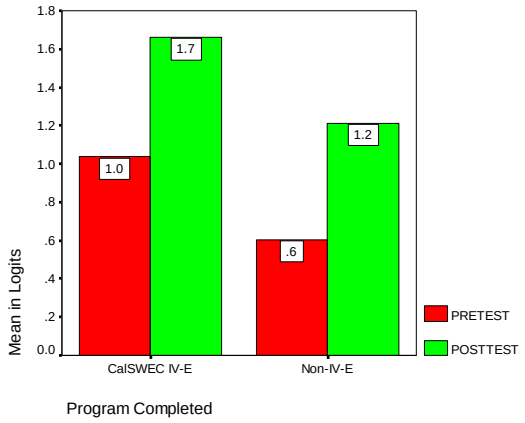
Test Versions: 1.0, 1.25, 1.26.

Date Range: Classes conducted between January 2007 and December 2007.

N=800 complete pairs of pre- and post-tests

- Gains from pre- to post-test are statistically significant for both Title IV-E and non-IV-E participants.
- Title IV-E participants scored slightly higher at pre-test and post-test than non-Title IV-E participants; however, these differences are not statistically significant.

Pre-test–Post-test Differences by Title IV-E Participation Calendar Year 2008



Test Versions: 1.25, 1.26.

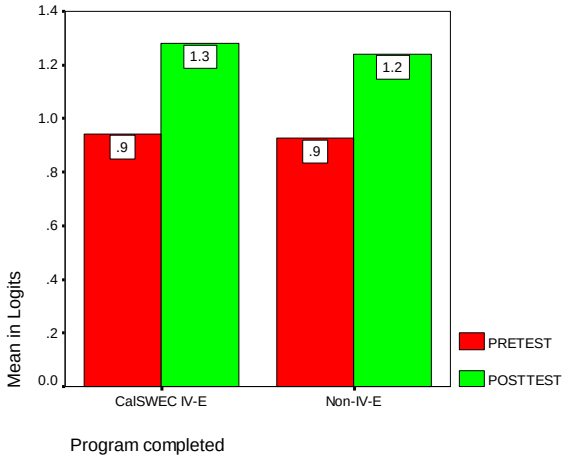
Date Range: Classes conducted between January 2008 and December 2008.

N=743 complete pairs of pre- and post-tests

- Gains from pre- to post-test are statistically significant for both Title IV-E and non-IV-E participants.
- Title IV-E participants scored higher at pre-test and post-test than non-Title IV-E participants. These differences are statistically significant.

*CURRICULUM TOPIC: CHILD AND YOUTH DEVELOPMENT

Pre-test–Post-test Differences by Title IV-E Participation Calendar Year 2007



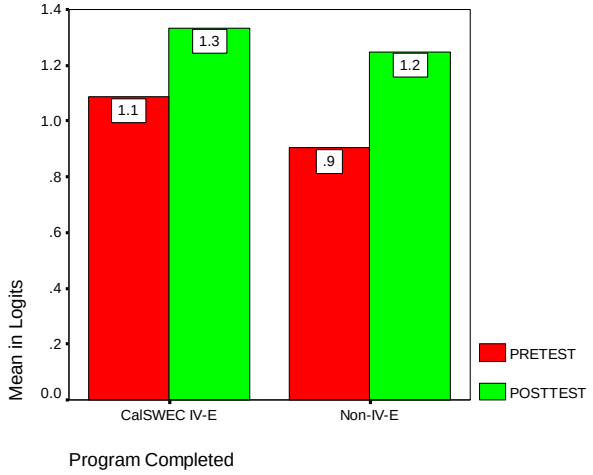
Test Versions: 1.0.

Date Range: Classes conducted between January 2007 and December 2007.

N=387 complete pairs of pre- and post-tests

- Gains from pre- to post-test are statistically significant for both Title IV-E and non-IV-E participants.
- Title IV-E participants scored at the same level as non-Title IV-E participants on the pre-test and slightly higher on the post-test. Post-test differences are not statistically significant.

Pre-test–Post-test Differences by Title IV-E Participation Calendar Year 2008



Test Versions: 1.0, 1.05

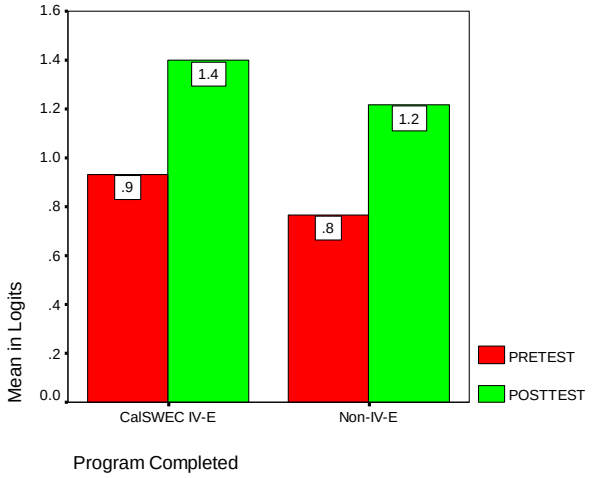
Date Range: Classes conducted between January 2008 and December 2008.

N=704 complete pairs of pre- and post-tests

- Gains from pre- to post-test are statistically significant for both Title IV-E and non-IV-E participants.
- Title IV-E participants scored significantly higher at pre-test non-Title IV-E.
- Title IV-E participants scored slightly higher at post-test; however, this difference is not statistically significant.

*CURRICULUM TOPIC: PLACEMENT AND PERMANENCY

Pre-test–Post-test Differences by Title IV-E Participation
Calendar Year 2007



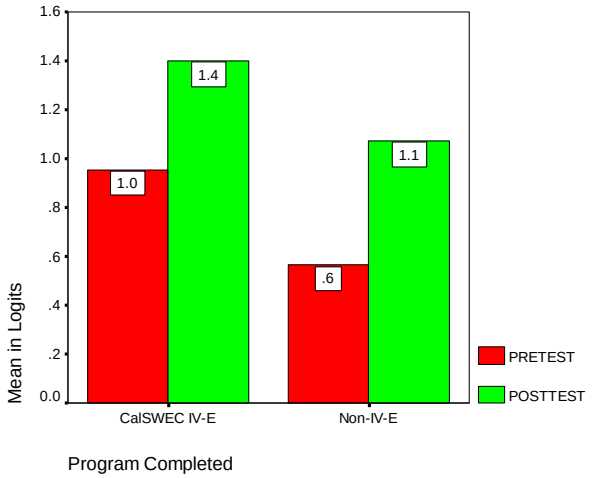
Test Versions: 1.0, 1.1.

Date Range: Classes conducted between January 2007 and December 2007.

N=878 complete pairs of pre- and post-tests

- Gains from pre- to post-test are statistically significant for both Title IV-E and non-IV-E participants.
- Title IV-E participants scored slightly higher at pre-test and post-test than non-Title IV-E participants; however, these differences were not statistically significant.

Pre-test–Post-test Differences by Title IV-E Participation
Calendar Year 2008



Test Versions: 1.1, 1.15.

Date Range: Classes conducted between January 2008 and December 2008.

N=742 complete pairs of pre- and post-tests

- Gains from pre- to post-test are statistically significant for both Title IV-E and non-IV-E participants.
- Title IV-E participants scored significantly higher at pre-test and post-test than non-Title IV-E participants.

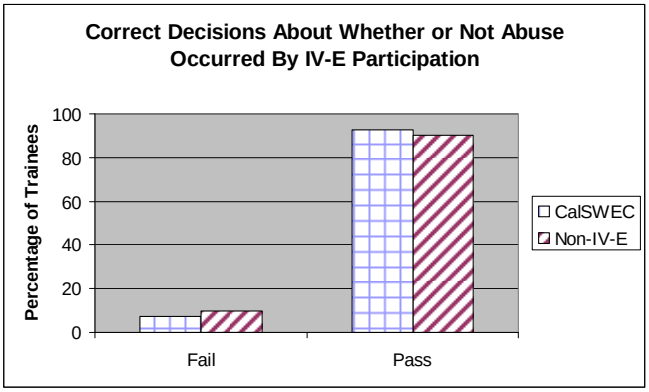
*CURRICULUM TOPIC: CHILD MALTREATMENT IDENTIFICATION, PARTS I AND II

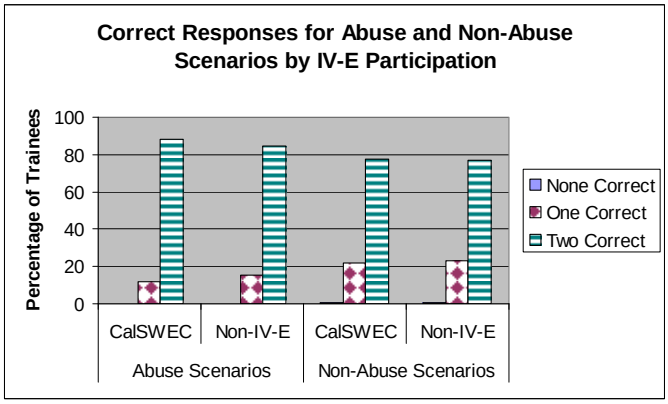
The strategic plan identified the need to evaluate the key skill area of child maltreatment identification with an embedded skills assessment. Embedded assessments build on activities that are already part of the training day both to provide evaluation data and reinforce key learning through additional practice and feedback.

The embedded evaluations for Child Maltreatment Identification, Parts I and II both require participants to analyze potential indicators of abuse in four different scenarios and make overall decisions about whether or not the children described have been victims of physical abuse (in module 1) or sexual abuse (in module II). In Part A, trainees must decide whether each of a set of individual elements that can indicate abuse is present in the scenario, is not present, or requires more information to make a decision. In Part B, trainees must make an overall decision about whether or not abuse has occurred, and in Part C they must indicate the three most important elements in making their decision. These are then given credit if they match those identified as most important by an expert group.

Results are presented here for the overall decision regarding abuse (Part B) for CMI 1 and CMI II for the most recent version of each:

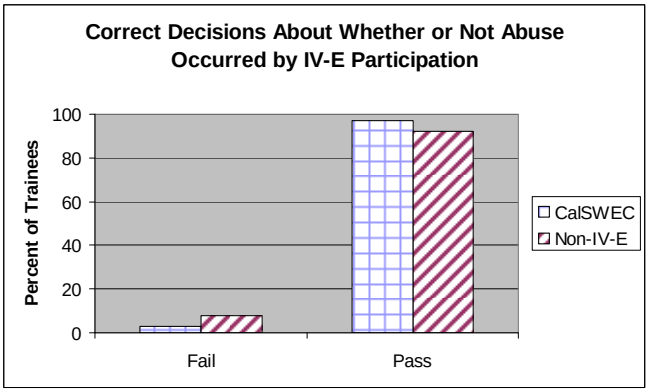
CMI I

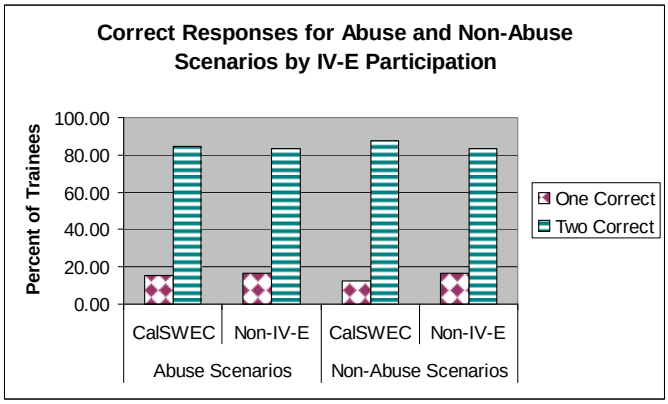




- Over 90% of trainees passed Part B with a minimum of 3 correct[15]: 92.7% of CalSWEC Title IV-E students, and 90.3% of non-Title IV-E students. These differences were not statistically significant.
- Trainees as a whole did slightly better when the decision was abuse than when it was not abuse, and this result may be due to a slight skew in the results. More specifically, trainees did especially well on one of the abuse scenarios where the perpetrator acknowledged hitting the child with a belt, which may have skewed somewhat the overall results. Trainees also had difficulty on one of the non-abuse scenarios in separating concerns over abuse from concerns over possible neglect.
- Differences in number correct associated with Title IV-E participation did not reach statistical significance for either abuse or non-abuse scenarios.

CMI II





- Over 92% of trainees passed Part B with a minimum of 3 correct[16]: 96.9% of CalSWEC Title IV-E students, and 92.1% of non-Title IV-E students. These differences were not statistically significant.
- Differences in number correct associated with Title IV-E participation did not reach statistical significance for either abuse or non-abuse scenarios.

---

[15] Graph based on 1807 responses to the latest versions (1.1. and 1.2) of the evaluation. All regions are represented.

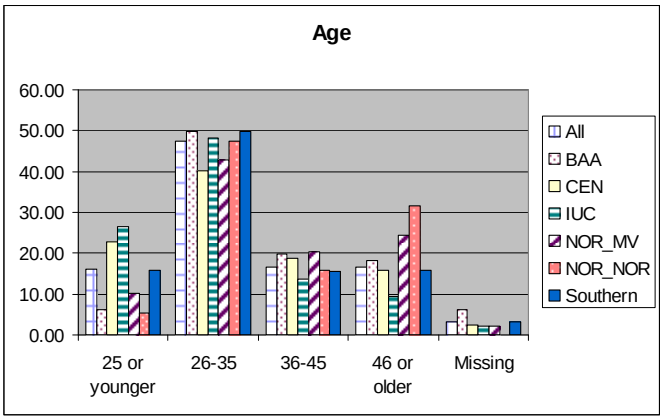
[16] Graph based on 425 responses to version 1.05 of the evaluation. All regions are represented except the Northern Academy for which no data were received.

- RESULTS FOR SELECT DEMOGRAPHIC INFORMATION -

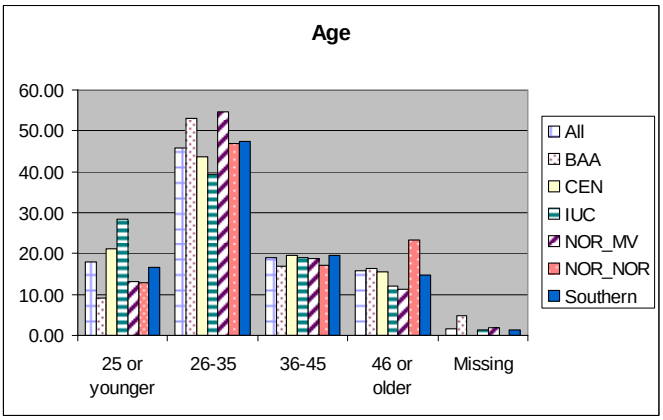
Trainee Demographic Profiles:

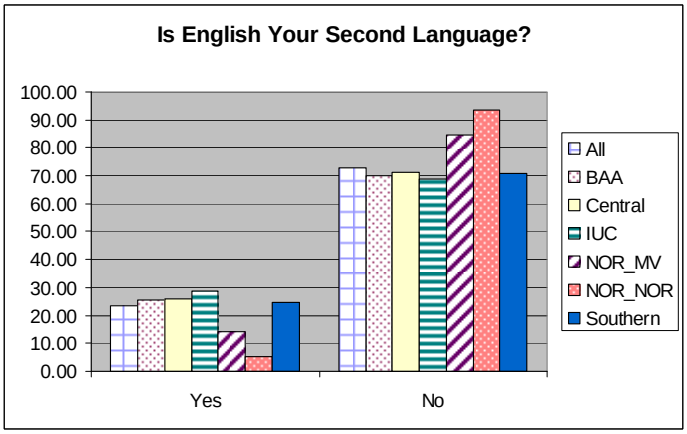
Trainee profiles are shown in the following tables for the key variables of age, race, ESL status, child welfare experience prior to their current position, educational level and Title IV-E participation, statewide and by region. Bars represent the percentage of trainees in each category.

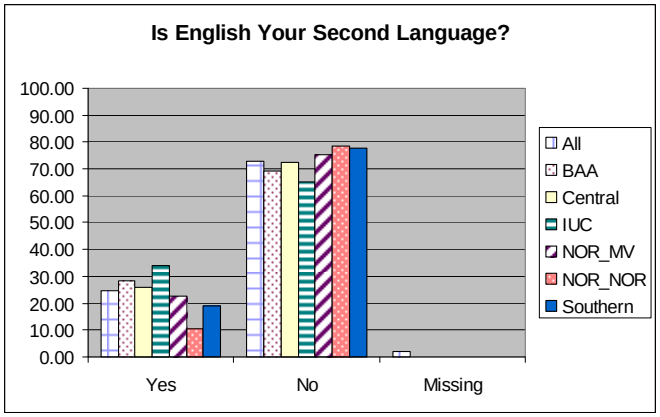
Demographics 2007



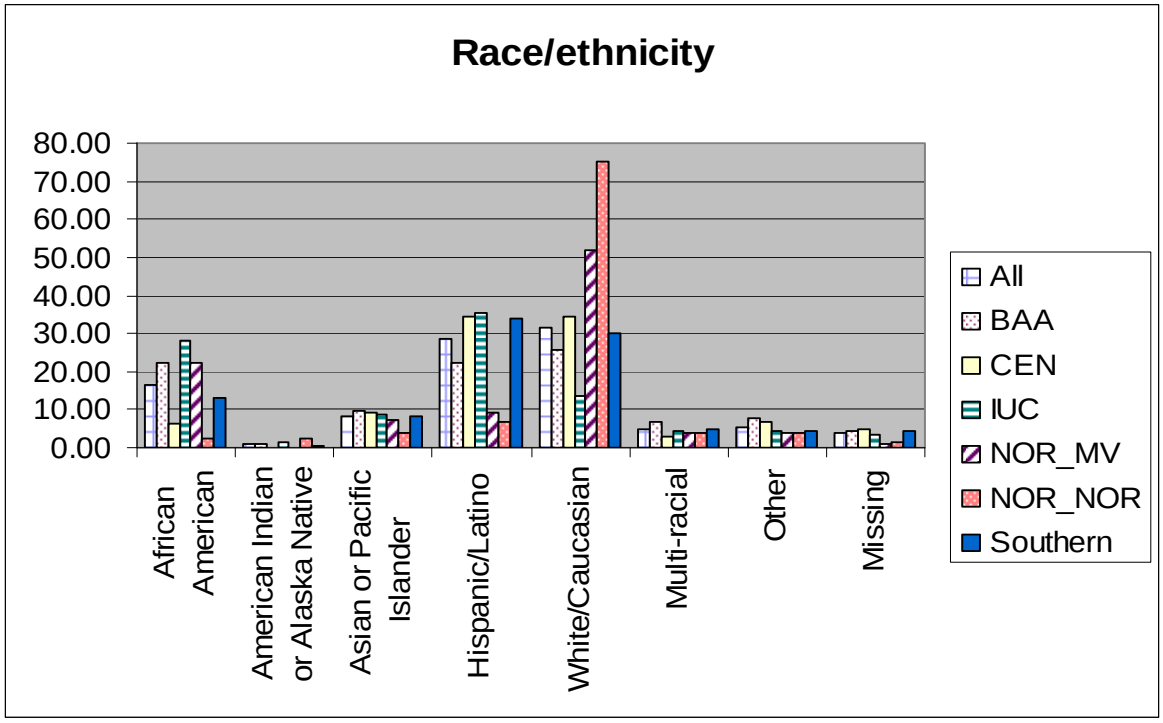
Demographics 2008



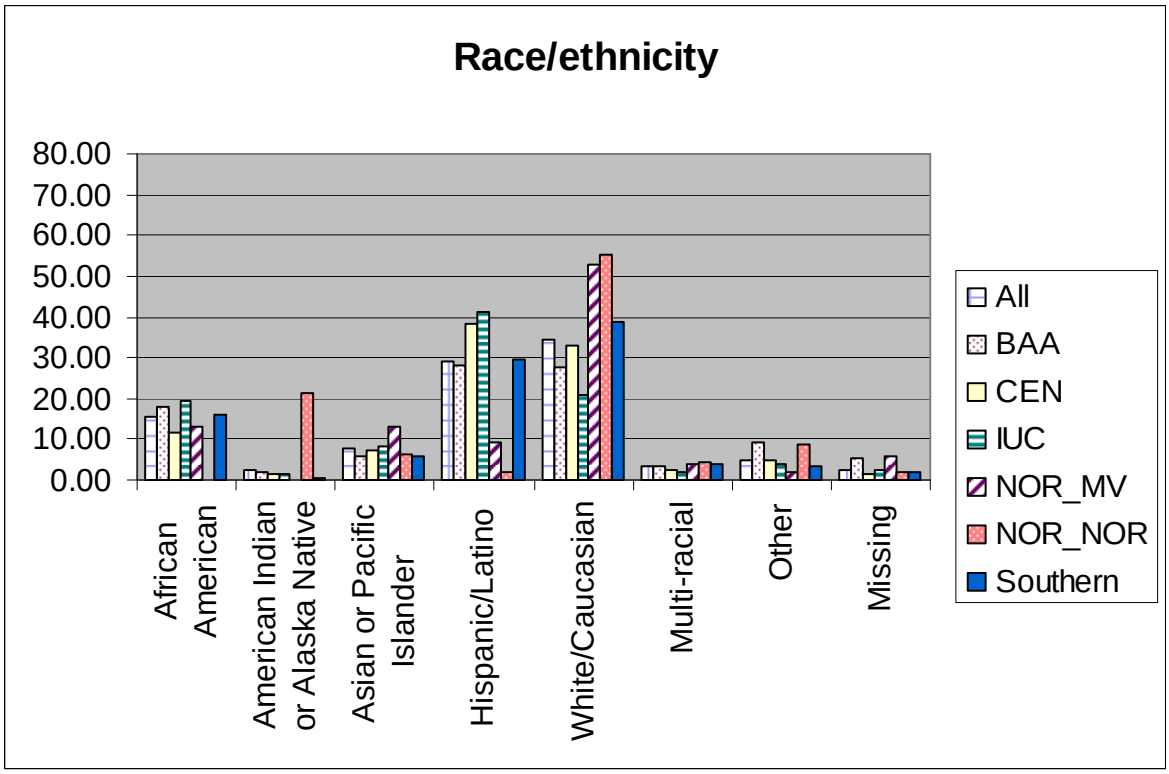




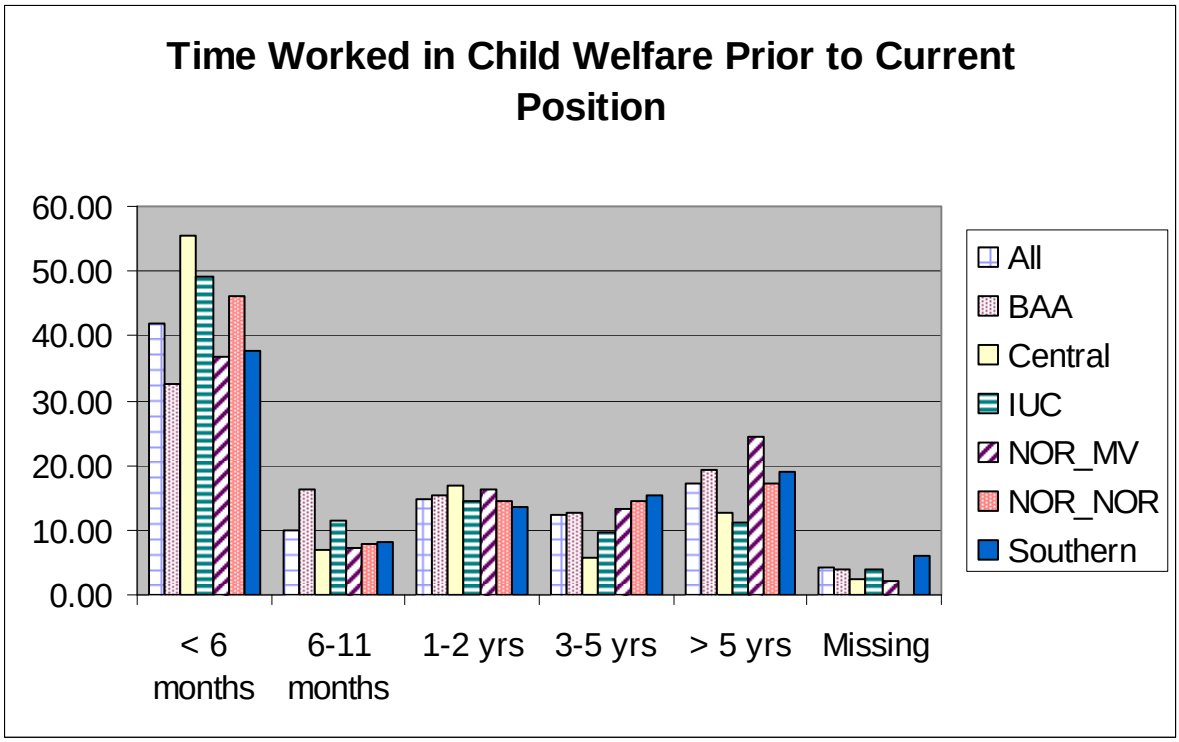
2007


Race/ethnicity
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
African American
American Indian or Alaska Native
Asian or Pacific Islander
Hispanic/Latino
White/Caucasian
Multi-racial
Other
Missing
All
BAA
CEN
IUC
NOR_MV
NOR_NOR
Southern

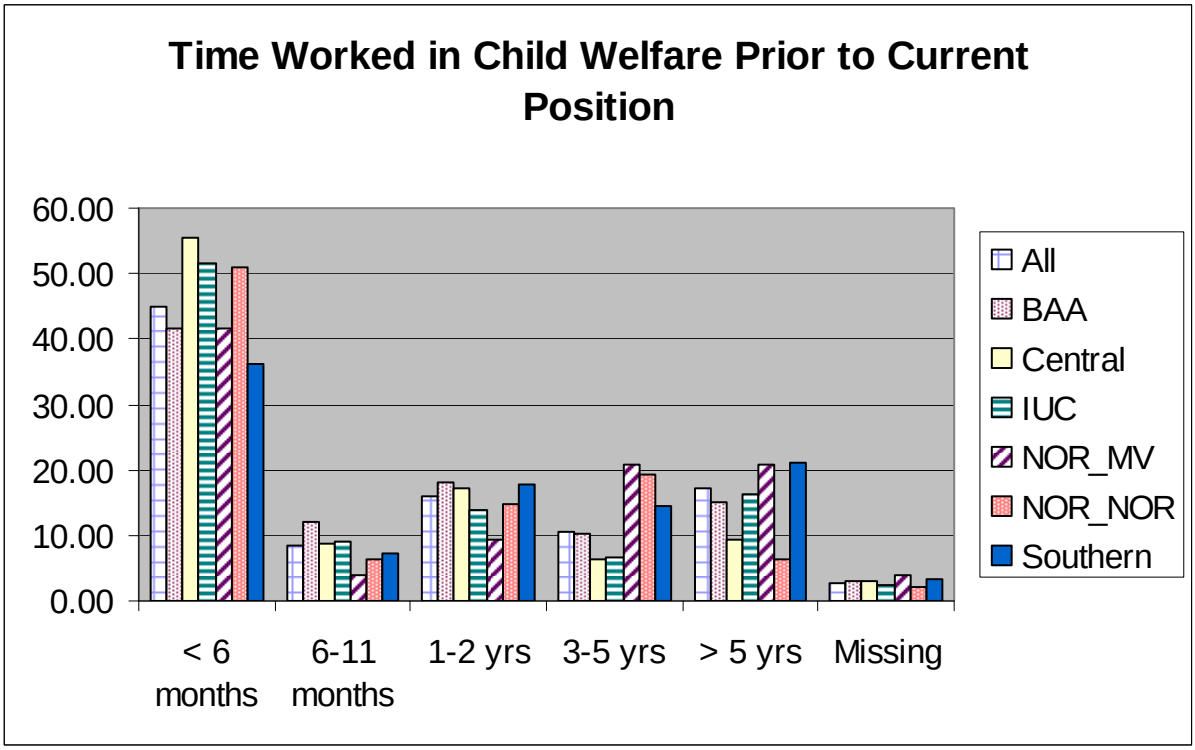
2008


Race/ethnicity
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
African American
American Indian or Alaska Native
Asian or Pacific Islander
Hispanic/Latino
White/Caucasian
Multi-racial
Other
Missing
All
BAA
CEN
IUC
NOR_MV
NOR_NOR
Southern

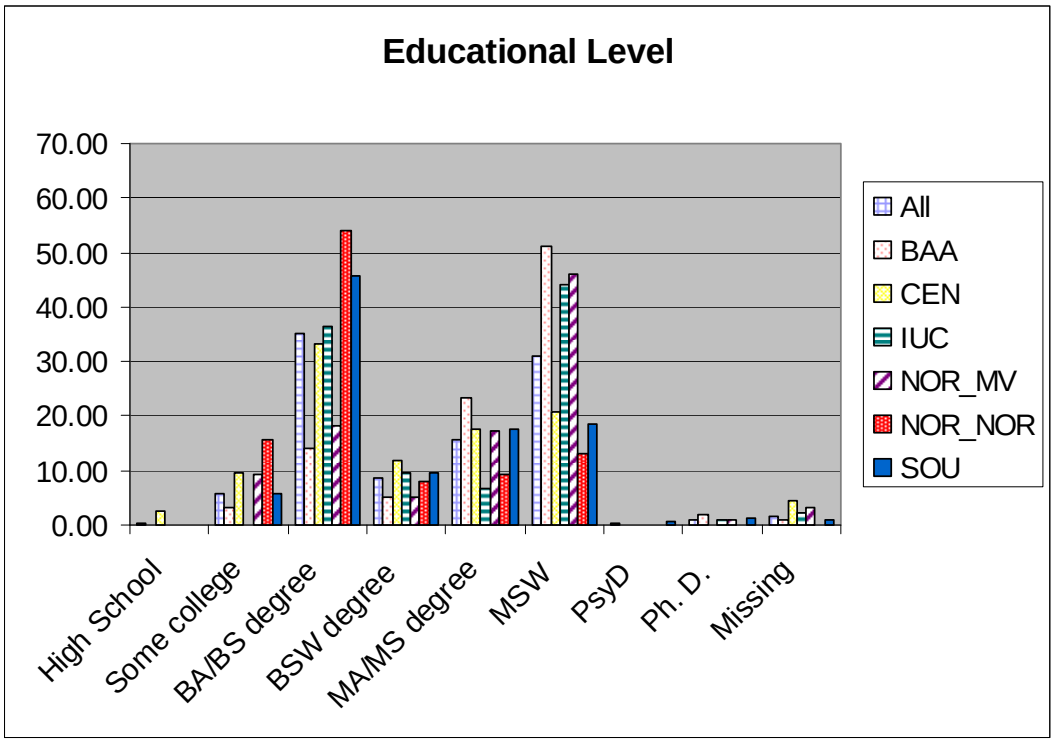
Education and Experience 2007



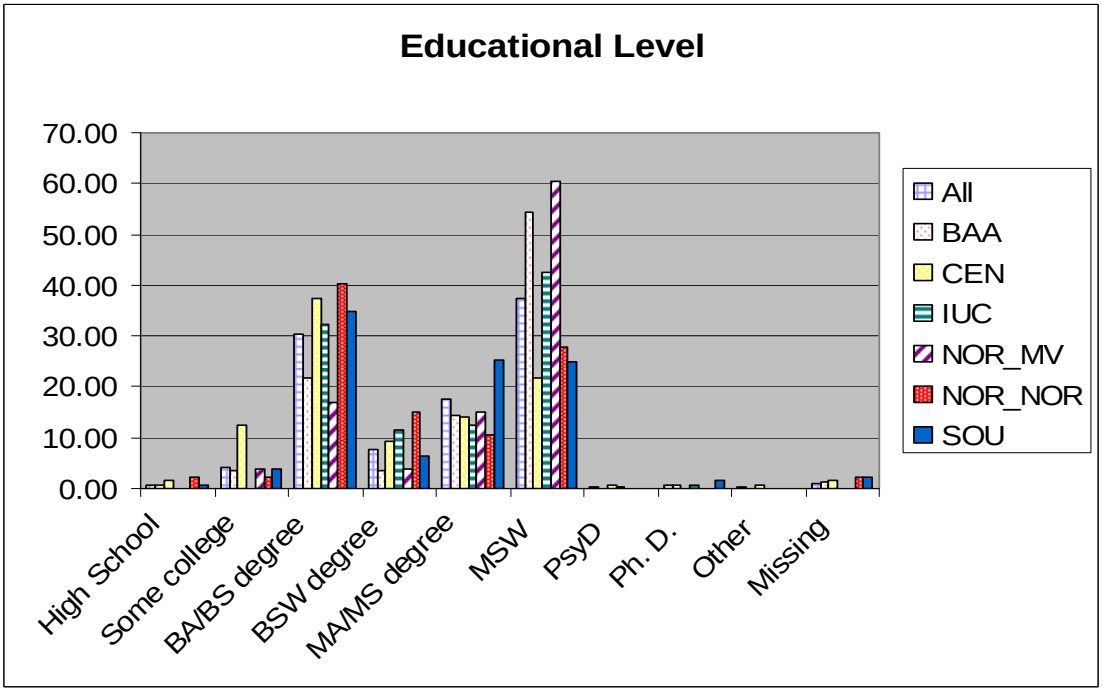
Education and Experience 2008



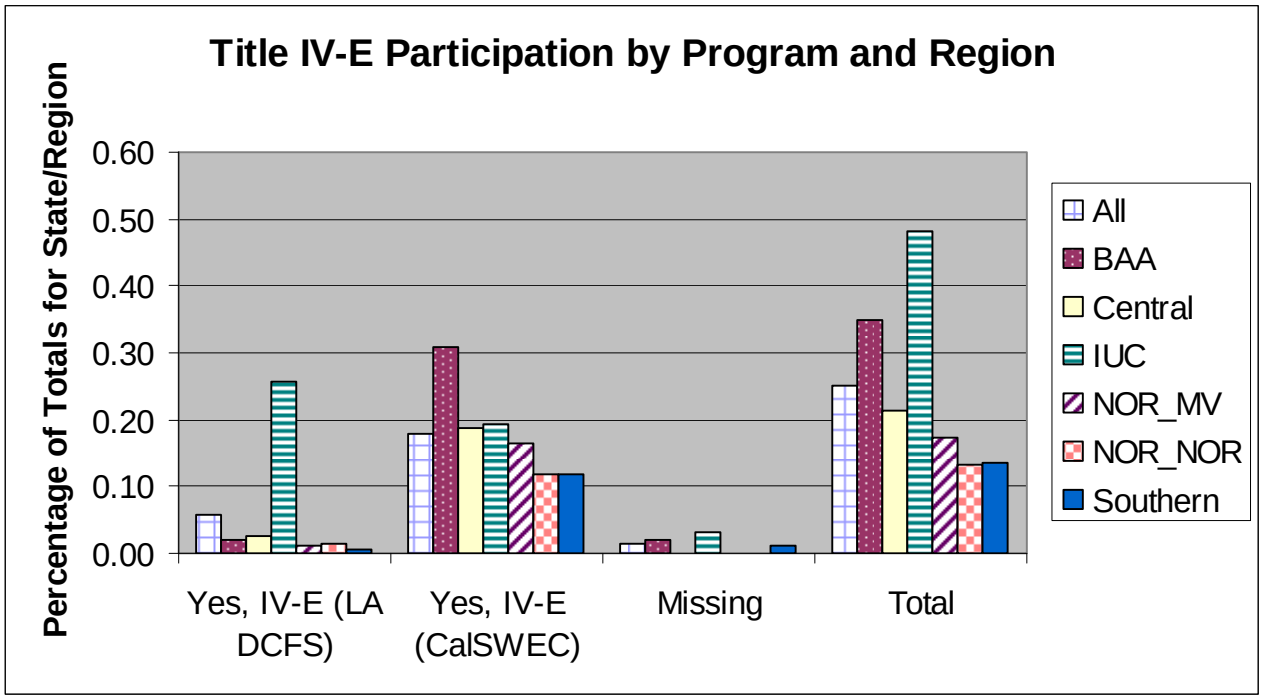
Education and Experience 2007



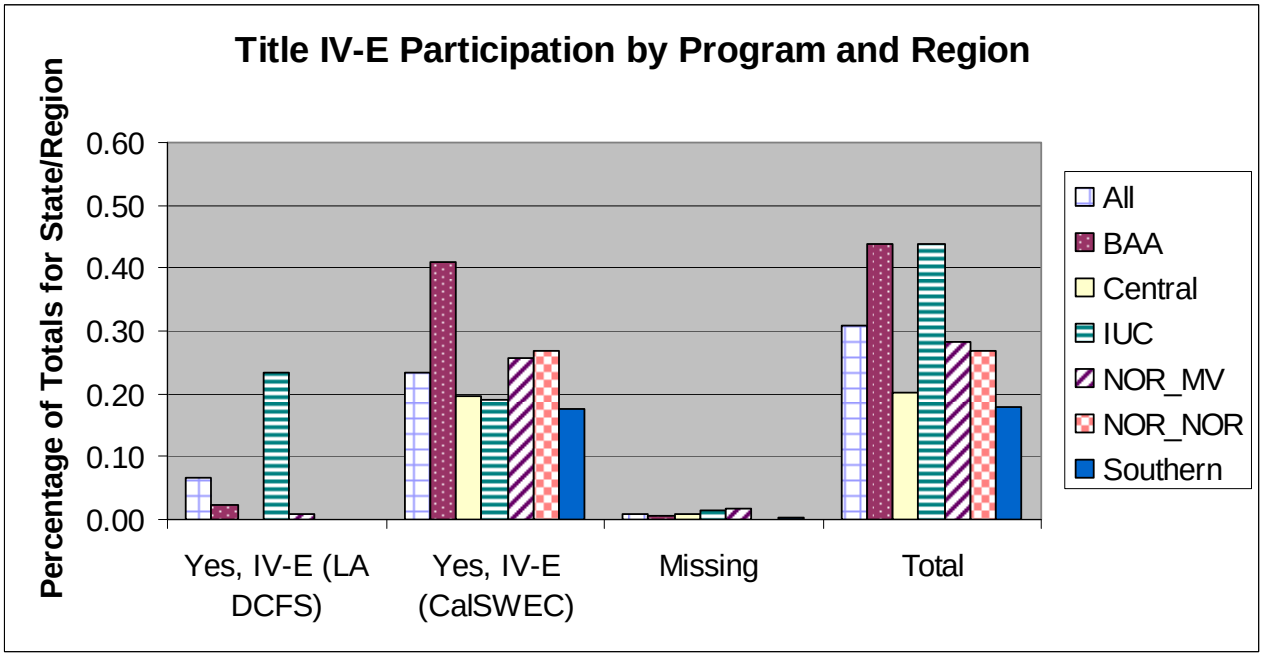
Education and Experience 2008



Title IV-E Participation 2007



Title IV-E Participation 2008



Appendix N: Sample Summary of Performance on Embedded Evaluations

CALIFORNIA SOCIAL WORK EDUCATION CENTER
UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF SOCIAL WELFARE

Statewide Report: Child Maltreatment Identification, Part 1
- March 2009 -

Overall Results

Figure 1

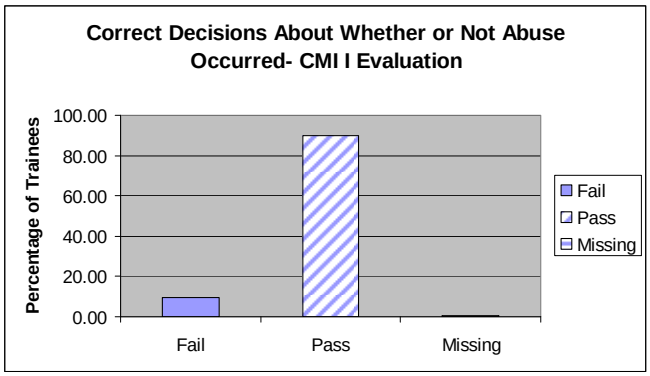
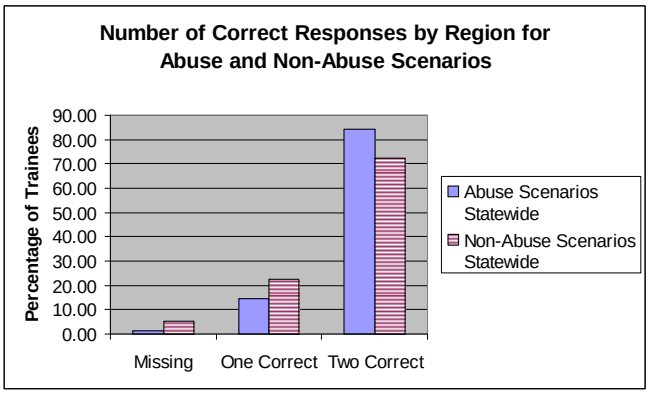


Figure 2



Data Included:
Date Range: February 2007 through December 2008
Academies represented[17]: BAA (N=240), CEN (N=263), IUC (N=485), NOR (N=298), SOU (N=521) (N=1807)

Curriculum version(s): 1.1 and 1.2
Answer sheet version: 1.1 and 1.2
Answer sheet versions 1.1 and 1.2 presented the Toby scenario first, rather than last as in earlier versions. Version 1.1 was used in a format that allowed an extended time period for completing the evaluation (2 hours as opposed to 90 minutes). Trainers later indicated that the extended timeframe was not needed and version 1.2 was used with a return to the 90 minute timeframe. Results were collapsed for this report after

[17] During this period the IUC used v 1.05 of the answer sheet in which the Toby scenario was presented last. These data have not been collapsed since the newer versions presented the scenarios in a different order (Toby first).

examination of error patterns and percentages of missing data did not indicate any substantial differences based on amount of time allowed for the evaluation.

Scoring: Preliminary criteria for adequate performance were established by the CDOG+ Team for CMI1 as follows:

- To "pass" part A the trainee needs to have 9 of the 13 items for each scenario correct
- To "pass" part B the trainee must have the correct decisions for 3 out of 4 scenarios
- To "pass" part C the trainee has to correctly identify 3 of the correct options.

The combination of scenario specific and overall test criteria has proven to be confusing and revisions to these criteria are currently under discussion. Thus, for this report, overall scores are presented for part B only. The first graph above shows the percentage of participants who pass part B with at least 3 out of 4 correct. Figure 2 shows correct answers by type of scenario (abuse or non-abuse). Bars represent the percentage of participants within each type of scenario who got only one decision correct compared to those who got both correct.

Results:

- Once again overall performance was strong on decision making in part B with 89.8% of trainees statewide passing, using a criterion of 3 out of 4 correct decisions.
  - Percentages correct by scenario were as follows:
    - Toby 82.5%
    - Miko 88.3%
    - Crystal 94.9%
    - D'Shan 86.3%
- There continued to be a higher percentage of correct decisions for Toby when that scenario appeared in the first position. As in the previous analyses, percentage correct for D'Shan, which is now in the last position, did not drop substantially. This suggests that time is not the only factor that has accounted for a lower percentage of correct decisions for Toby.
- Percentage of participants who made two correct decisions on Part B, continued to be higher for abuse scenarios (82.4%) compared to non –abuse scenarios (72.3%).
- Percentages of those who made only one correct decision was higher for non-abuse scenarios (22.4%) compared to abuse scenarios (14.4%), as was the percentage of missing data (5.3% compared to 1.3%). Higher percentages of missing data occurred primarily on the D'Shan scenario that was administered last. This suggests that time may have affected these comparisons; however, the effect is relatively small and the percentage of participants who passed the D'Shan scenario is still in line with percentages for other scenarios.
- The following elements were identified as potential concerns on part A: Error patterns have been very consistent over time despite small changes to the scenarios and wording of the elements, and merit further attention (e.g. discussions with trainers or a group of participants). Specific problem items were:
  - Toby: items 3, 4, and 9.
  - D'Shan: items 2, 3 and 10.
  - Crystal: items 1, 5, 7, 8, 10 and 11.
  - Miko: item 9.

- Item analysis showed that the following items had negative discrimination values[18]:
  - Toby: item 7, Part A
  - D'Shan: Items 8 and 11, Part A
  - Crystal: Item 8 Part A

- Participants continued to be better able to match the key decision making elements identified by the curriculum developers in part C, on abuse scenarios compared to non-abuse scenarios.

Details of performance on specific elements are shown in the tables below.

[18] Negative discrimination in testing refers to a situation in which high scorers miss the item or low scorers get the item correct more often than expected. It can occur when an item is confusing, has more than one potential correct answer, has no clear correct answer, or is at an extreme of difficulty (either too hard or too easy) which can trigger errors due to inattention or guessing.

TOBY CASE SCENARIO

Part A Item Analysis (Shaded percentages are correct)

ELEMENT	YES	NO	Need more information
1) Location/scene of incident	**16.9%**	**76.4%**	**5.3%**
2) Location of injuries on the body	**30.5%**	**67.1%**	**1.7%**
3) Type(s) of injury	**38.3%**	**58.2%**	**2.7%**
4) Severity/extent of current injuries	**44.5%**	**50.6%**	**3.2%**
5) Frequency of injuries over time	**70.4%**	**24.7%**	**3.1%**
6) Explanation of injuries	**19.0%**	**74.9%**	**4.2%**
7) Child's overall appearance	**69.7%**	**23.1%**	**5.1%**
8) Chronological age of child	**11.3%**	**87.2%**	**.7%**
9) Developmental abilities of child	**18.6%**	**65.5%**	**14.4%**
10) History of unreported maltreatment -of these children by any caregivers -by these caregivers toward any children	**13.3%**	**70.4%**	**15.1%**
11) History of CWS involvement: -with these children and/or -with these caregivers	**2.1%**	**85.2%**	**12.2%**
12) Current caregivers' history of own childhood maltreatment	**1.0%**	**20.8%**	**77.5%**
13) Current caregivers' substance abuse	**.2%**	**20.4%**	**78.8%**

Underlined and italicized items had percentages correct of less than 2/3rds. They might need a second look to be sure that the scenario itself and/or the teaching of the decision point is clear.

1) Location/scene of incident—These injuries appear to have occurred in the home and at the skate park. The skate park location does not suggest abuse. The home location does not affect whether these injuries meet the definition of abuse one way or the other.
2) Location of injuries on body—Locations of some of the injuries are more consistent with accident than with abuse (little toe, knees, elbows, and especially shins). The location of the other injuries (arm, face) could be either. "A chipped tooth, while located in the head, is not necessarily a problematic location for injury."
3) *Type of injury*—Toby has several types of injuries (broken bone, chipped tooth, bruises), none of which is more associated with infliction than with accident.
4) *Severity/extent of current injuries* —A broken bone is a severe injury.
5) Frequency of injuries over time—Toby has many injuries over time.
6) Explanation of injuries—The explanation for the injuries are all consistent with the injuries themselves. Toby, his mother and brother all agree on the explanations and neither child appears to be coerced.
7) Child's overall appearance— Child appears overweight and often is disheveled: his clothes are torn and sometimes dirty and his long hair is often messy. In addition, he has multiple bruises on the shins and evidence of skinned elbows and knees. It should be noted that some of these characteristics can be understood within a cultural context (e.g., 'skateboarder culture').
8) Chronological age of child —This child is older.

9) *Developmental abilities of the child*—This child appears to be developmentally on target.
10) History of unreported maltreatment:-of these children by any caregivers,-by these caregivers toward any children —There is no known history.
11) History of CWS involvement:-with these children and/or-with these caregivers —There is no CWS history.
12) Current caregivers' history of own childhood maltreatment—There is no information on this yet.
13) Current caregivers' substance abuse—There is no information about this yet.

B. With the information currently available and based on your assessment of the elements what MOST LIKELY will be your decision: (Shaded item is correct)

1)	Child maltreatment occurred	16.1%
2)	Child maltreatment did not occur	82.5%

C. Of the following elements, which THREE were MOST significant to you in making your decision in Question B? Any three of the shaded options is considered correct. Percentages are percentages of participants who checked that option. Low percentages in a shaded area may signal that the importance of that factor was missed. However, it may also just mean that other factors were viewed by most participants as more important so that factor didn't make the top three.

19.8%	1)	Location of incident
67.8%	2)	Location of injuries on the body
66.1%	3)	Type(s) of injury
14.7%	4)	Severity of injuries
17.5%	5)	Frequency of injuries over time
75.4%	6)	Explanation of injuries
10.8%	7)	Overall appearance of child
28.1%	8)	Chronological age of child
4.8%	9)	Developmental abilities of child
2.8%	10)	History of unreported maltreatment:-of these children by any caregivers,-by these caregivers toward any children
2.7%	11)	History of CWS involvement:-with these children and/or-with these caregivers
.3%	12)	Current caregivers' history of own childhood abuse
.1%	13)	Current caregivers' substance abuse

Toby does not appear to have been abused. The elements are:

- The types and locations of the injuries fit with the explanation of them and the fact that both children explain the injuries in the same way as their mother does and that neither child appears to be coerced.
- Toby appears to be in good health and is old enough to let others know if he is being hurt or a least to give signs that he is afraid. He does not appear to be afraid.

MIKO CASE SCENARIO

Part A Item Analysis (Shaded percentages are correct)

ELEMENT	YES	NO	Need more information
1) Location/scene of incident	**77.9%**	**18.9%**	**2.2%**
2) Location of injuries on the body	**97.1%**	**2.3%**	**.2%**
3) Type(s) of injury	**97.2%**	**1.7%**	**.4%**
4) Severity/extent of current injuries	**93.3%**	**4.9%**	**.9%**
5) Frequency of injuries over time	**4.9%**	**87.2%**	**6.6%**
6) Explanation of injuries	**83.3%**	**12.0%**	**3.5%**
7) Child's overall appearance	**10.9%**	**85.6%**	**2.1%**
8) Chronological age of child	**94.4%**	**4.5%**	**.6%**
9) Developmental abilities of child	**37.8%**	**53.2%**	**7.7%**
10) History of unreported maltreatment -of these children by any caregivers -by these caregivers toward any children	**1.2%**	**83.3%**	**13.7%**
11) History of CWS involvement: -with these children and/or -with these caregivers	**1.6%**	**90.5%**	**7.4%**
12) Current caregivers' history of own childhood maltreatment	**.6%**	**21.5%**	**77.2%**
13) Current caregivers' substance abuse	**.6%**	**22.5%**	**76.4%**

Underlined and italicized items had percentages correct of less than approximately 2/3rds. They might need a second look to be sure that the scenario itself and/or the teaching of the decision point is clear.

1) Location/scene of incident—The couch in the family's home is a concern because it is a dangerous place to leave an infant who is learning to roll over unattended. However, the fact that the couch is two feet off the ground and on a rug suggests that any injury sustained would likely be minor unless the falling infant hit something besides the floor first.
2) Severity/extent of current injuries —The leg of a young infant is not a likely locale for an accidental injury because the baby cannot crawl.
3) Type of injury—The baby has a spiral fracture, an injury which occurs only if the baby's leg is twisted (as explained to the father by the nurse). This is highly unlikely to occur by accident.
4) Severity of injury—A broken leg is a serious injury.
5) Frequency of injuries over time—This is the first known injury of this baby.
6) Explanation of injuries—Babies who roll off of couches do not twist their legs. This baby's leg was twisted. The explanation does not fit with the facts.
7) Child's overall appearance—Child appears healthy except for this injury.
8) Chronological age of child —This child is very young and thus vulnerable.
9) *Developmental abilities of the child*—This infant's age indicates that her developmental abilities are on track by age.

10) History of unreported maltreatment:-of these children by any caregivers,-by these caregivers toward any children —There is no evidence or concern that prior unreported maltreatment of Miko has occurred or that her parents have maltreated another child.
11) History of CWS involvement:-with these children and/or-with these caregivers —There is no CWS history.
12) Current caregivers' history of own childhood maltreatment—There is no information on this yet.
13) Current caregivers' substance abuse—There is no information about this yet.

B. With the information currently available and based on your assessment of the elements what MOST LIKELY will be your decision: (Shaded item is correct.)

1)	Child maltreatment occurred	88.3%
2)	Child maltreatment did not occur	10.0%

C. Of the following elements, which THREE were MOST significant to you in making your decision in Question B? Any three of the shaded options is considered correct. Percentages are percentages of participants who checked that option. Low percentages in a shaded area may signal that the importance of that factor was missed. However, it may also just mean that other factors were viewed by most participants as more important so that factor didn't make the top three.

34.1%	1)	Location/scene of incident
45.7%	2)	Location of injuries on the body
74.8%	3)	Type(s) of injury
46.2%	4)	Severity of injuries
2.2%	5)	Frequency of injuries over time
61.6%	6)	Explanation of injuries
4.8%	7)	Overall appearance of child
43.2%	8)	Chronological age of child
8.0%	9)	Developmental abilities of child
1.0%	10)	History of unreported maltreatment:-of these children by any caregivers,-by these caregivers toward any children
1.0%	11)	History of CWS involvement:-with these children and/or-with these caregivers
.2%	12)	Current caregivers' history of own childhood abuse
.1%	13)	Current caregivers' substance abuse

It is likely that Miko was abused. The elements are:

- The location of the injury on the body (the leg) is not a common spot of an accidental injury in a child.
- The type of injury is caused by a twisting motion, which does not happen accidentally in an infant.
- The severity of the injury is high.
- The father's explanation is not consistent with the nature of the injury.
- This injury would not have occurred at the location (i.e., a couch two feet off of a carpeted floor).
- The age of the child makes her extremely vulnerable.

CRYSTAL CASE SCENARIO

Part A Item Analysis (Shaded percentages are correct)

ELEMENT	YES	NO	Need more information
1) Location/scene of incident	**29.1%**	**65.1%**	**4.0%**
2) Location of injuries on the body	**97.2%**	**2.1%**	**.1%**
3) Type(s) of injury	**97.7%**	**1.8%**	**.1%**
4) Severity/extent of current injuries	**71.6%**	**25.7%**	**1.8%**
5) Frequency of injuries over time	**59.6%**	**27.3%**	**11.7%**
6) Explanation of injuries	**91.8%**	**6.7%**	**.3%**
7) Child's overall appearance	**16.4%**	**60.9%**	**21.0%**
8) Chronological age of child	**32.5%**	**65.4%**	**.6%**
9) Developmental abilities of child	**93.2%**	**4.7%**	**.8%**
10) History of unreported maltreatment -of these children by any caregivers -by these caregivers toward any children*	**46.1%**	**36.9%**	**15.4%**
11) History of CWS involvement: -with these children and/or -with these caregivers	**49.2%**	**40.5%**	**9.5%**
12) Current caregivers' history of own childhood maltreatment	**4.2%**	**19.3%**	**75.5%**
13) Current caregivers' substance abuse	**2.8%**	**22.3%**	**73.9%**

* Missing responses over 5%

Underlined and italicized items had percentages correct of less than approximately 2/3rds. They might need a second look to be sure that the scenario itself and/or the teaching of the decision point is clear.

1) *Location/scene of incident* —These injuries happened at home and there is nothing about this location that suggests that this has an impact on whether the injuries meet the definition of abuse or not.
2) Location of injuries on body—The injuries are on multiple spots on the back of her legs—an unlikely place for accidental multiple injuries.
3) Type of injury—The injuries are bruises and they are due to whipping with a belt.
4) Severity/extent of current injuries —While the bruises are not severe, the fact that there are multiple bruises means that she was struck repeatedly, which increases severity.
5) *Frequency of injuries over time*—Grandfather says he has hit her before this incident in the same way (with a belt)
6) Explanation of injuries—Grandfather acknowledges that he hit her with a belt and this left bruises.
7) *Child's overall appearance*—No information is provided.
8) *Chronological age of child* —The child is older.
9) Developmental abilities of the child—Crystal is developmentally delayed.
10) *History of unreported maltreatment:*-of these children by any caregivers,-by these caregivers toward any children —Crystal was neglected as a baby by her mother. Her grandfather acknowledges that he hit her with a belt on two occasions.

11) *History of CWS involvement:-*with these children and/or-with these caregivers —There is a CWS history—Crystal was neglected by her mother as a young child. There is no known CWS history with the grandparents who are now raising her.
12) Current caregivers' history of own childhood maltreatment—There is no information on this yet.
13) Current caregivers' substance abuse—There is no information about this yet.

B. With the information currently available and based on your assessment of the elements what MOST LIKELY will be your decision: (Shaded item is correct.)

1) Child maltreatment occurred 94.9%
2) Child maltreatment did not occur 3.0%

C. Of the following elements, which THREE were MOST significant to you in making your decision in Question B? Any three of the shaded options is considered correct. Percentages are percentages of participants who checked that option. Low percentages in a shaded area may signal that the importance of that factor was missed. However, it may also just mean that other factors were viewed by most participants as more important so that factor didn't make the top three.

4.4%	1)	Location of incident
62.5%	2)	Location of injuries on the body
65.1%	3)	Type(s) of injury
23.3%	4)	Severity of injuries
19.0%	5)	Frequency of injuries over time
71.8%	6)	Explanation of injuries
.9%	7)	Overall appearance of child
5.1%	8)	Chronological age of child
57.3%	9)	Developmental abilities of child
4.2%	10)	History of unreported maltreatment:-of these children by any caregivers,-by these caregivers toward any children
5.0%	11)	History of CWS involvement:-with these children and/or-with these caregivers
.4%	12)	Current caregivers' history of own childhood abuse
.1%	13)	Current caregivers' substance abuse

Crystal was abused. The elements are:

- The location of the bruises are on the back of the legs.
- The injuries are bruises resulting from a whipping with a belt, which grandfather acknowledges.
- The child is developmentally delayed which likely makes it harder for her to protect herself. (This is a good point to clarify the importance of differentiating the elements of chronological age from developmental level).

D'SHAN CASE SCENARIO

Part A Item Analysis (Shaded percentages are correct)

ELEMENT	YES	NO	Need more information
1) Location/scene of incident	**19.6%**	**71.9%**	**5.1%**
2) Location of injuries on the body	**55.8%**	**40.2%**	**1.8%**
3) Type(s) of injury	**50.0%**	**45.3%**	**2.3%**
4) Severity/extent of current injuries	**27.6%**	**66.7%**	**3.1%**
5) Frequency of injuries over time	**16.5%**	**70.7%**	**9.7%**
6) Explanation of injuries	**21.6%**	**72.7%**	**3.2%**
7) Child's overall appearance	**18.1%**	**74.0%**	**4.6%**
8) Chronological age of child	**73.8%**	**23.5%**	**.4%**
9) Developmental abilities of child	**16.1%**	**70.2%**	**10.5%**
10) History of unreported maltreatment -of these children by any caregivers -by these caregivers toward any children	**24.0%**	**57.8%**	**14.4%**
11) History of CWS involvement: -with these children and/or -with these caregivers	**78.3%**	**15.7%**	**3.7%**
12) Current caregivers' history of own childhood maltreatment	**2.3%**	**21.9%**	**73.5%**
13) Current caregivers' substance abuse	**1.4%**	**21.1%**	**75.0%**

Underlined and italicized items had percentages correct of less than approximately 2/3rds. They might need a second look to be sure that the scenario itself and/or the teaching of the decision point is clear.

1) Location/scene of incident —The location of the injuries appear to be in the home and there is no reason to think this affects the decision about whether this situation meets the definition of abuse in one way or another.
2) *Location of injuries on body*— Two of the injuries are on parts of the body that could be due to accident or have been inflicted (lower back and forehead). One location (shin) is highly associated with accidents only. Also, the mark on the forehead is close to the brain and could indicate some type of head trauma.
3) *Type of injury*—The two injuries are bruises. The third mark is a birthmark.
4) Severity/extent of current injuries —Neither of the injuries is severe.
5) Frequency of injuries over time—There are only two injuries from separate times. In addition, it is not possible to date bruises based on bruise coloration.
6) Explanation of injuries—Mother's explanation of injuries as accidental is consistent with the injuries and corroborated by her sister.
7) Child's overall appearance—Child appears healthy.
8) Chronological age of child —This child is young.
9) Developmental abilities of the child—This child appears on target developmentally
10) *History of unreported maltreatment:*-of these children by any caregivers,-by these caregivers toward any children Although trainees might mark "yes" for this element, there is no evidence or concern that prior unreported maltreatment of D'Shan has occurred or that his parents have maltreated another child.

11) History of CWS involvement:-with these children and/or-with these caregivers — There is a CWS history – one prior report based on D'Shan's father having left D'Shan alone at two months of age while responsible for babysitting him.
12) Current caregivers' history of own childhood maltreatment—There is no information on this yet.
13) Current caregivers' substance abuse—There is no information about this yet.

B. With the information currently available and based on your assessment of the elements what MOST LIKELY will be your decision: (Shaded item is correct.)

1)	Child maltreatment occurred	9.6%
2)	Child maltreatment did not occur	86.3%

C. Of the following elements, which THREE were MOST significant to you in making your decision in Question B? Any three of the shaded options is considered correct. Percentages are percentages of participants who checked that option. Low percentages in a shaded area may signal that the importance of that factor was missed. However, it may also just mean that other factors were viewed by most participants as more important so that factor didn't make the top three.

16.8%	1)	Location of incident
63.6%	2)	Location of injuries on the body
58.9%	3)	Type(s) of injury
12.2%	4)	Severity of injuries
5.3%	5)	Frequency of injuries over time
83.4%	6)	Explanation of injuries
15.9%	7)	Overall appearance of child
30.3%	8)	Chronological age of child
7.5%	9)	Developmental abilities of child
1.3%	10)	History of unreported maltreatment:-of these children by any caregivers,-by these caregivers toward any children
7.1%	11)	History of CWS involvement:-with these children and/or-with these caregivers
.3%	12)	Current caregivers' history of own childhood abuse
.2%	13)	Current caregivers'

D'Shan is likely not abused. He has only two injuries and one is almost certainly a result of an accident (bruise on shin). The other injury is explained adequately by the mother and her explanation is corroborated by her sister. The third mark is not an injury. There are no other indicators that make this situation suspicious.

Note that there are many areas where bias could influence decision making in this case. These include:

- Young age of mother.
- Mother is African American.
- Mother is annoyed—Could be seen as having an attitude and not cooperative.
- A crisis in mother's family led to use of the provider who reported to CWS (Mother's brother was in hospital being stitched up for injuries incurred as a result of having been beat up by peers. This led to grandmother being unable to take care of D'Shan.)
- Child's other parent once left him alone and CWS became involved.

Appendix O: Strategic Plan for Child Welfare Training Evaluation 2009–2012, Summary Matrix

Macro Evaluation Subcommittee Strategic Planning Process (Finalized October 2009): Matrix for Identifying Options and Prioritizing Efforts for the Next Three Years

The projects listed in the chart were suggested during the Macro Evaluation Subcommittee meeting July 22, 2009 and then finalized during the subsequent Macro Evaluation Subcommittee meeting (9/21/09) and conference call (10/19/09).

I. Horizontal Projects: Expanding/developing new evaluation projects at levels where current projects are underway, e.g., Level 3: Course level (formative) evaluation, Level 4: Knowledge and Level 5: Skills-during-training (embedded) evaluations.

DESCRIPTION OF POSSIBLE OPTIONS	RESOURCES NEEDED	PROJECTED TIMEFRAME
A. New knowledge tests for an existing or new courses **Course(s):** *Purpose:* To identify the degree of knowledge learning being achieved by trainees by the end of training and to then focus course revisions or delivery methods to help trainees master knowledge content. ***Recommendations:*** 1. Do not develop brand new knowledge tests at this time for Phase I (Big 7) topics that do not already have knowledge tests associated with them. 2. Keep knowledge tests for Phase I (Big 7) courses for which tests already exist, including test revisions. 3. Move toward more diagnostic pre- post-tests: a. Weight LOs for curricula b. Target a limited number of LOs for each pre-post knowledge test (i.e.,, don't test all LOs on evals; keep to 30 item tests but with items focused on specific LOs) c. Identify the LOs to be tested and develop more items as needed.		Start: FY09-10

DESCRIPTION OF POSSIBLE OPTIONS	RESOURCES NEEDED	PROJECTED TIMEFRAME
B. New skills-during-training assessments for existing or new courses. Content area/course(s): Neglect *Purpose:* To promote and assess trainee mastery of key skills during training. *Recommendations:* 1. Conduct needs assessment first to understand practice differences in assessing neglect. (e.g., Can we find a scenario that all counties agree is neglect? However, do not do a full needs assessment of counties' practices assessing neglect because this is highly resource intensive and budget is limited.) 2. Tweak existing scenario (Toby from CMI1) to create a neglect embedded evaluation via either pencil/paper or computer-assisted training tool within the current CMI1 curriculum. 3. Pilot in one region (Southern RTA volunteered) to determine if this modality for training and evaluation on this topic is applicable/useful for wider use.	- Computers (this may be an issue for counties) - Software (SIMS or Second Life; again, this may be an issue for counties)	Start: FY09-10

C. Revise the tools for formative evaluation of curricula:
Purpose: To ensure that course development meets curricula standards
 	
Recommendation: Revise formative evaluation tools:
- Priority: Revise observer's version of formative evaluation tools to include rubric on the tool itself. Use previous comments on formative evaluation tools to develop a rubric and ensure that rubric is included on the final version of the observer evaluation tool (Look at Northern RTA's scale for example.) Eliminate potential evaluation of trainers (keeps focus separate so we know content is being evaluated when that's the focus vs. trainers being evaluated when that's not the focus). Create separate tools for assessing:
  a. content
  b. methods of delivery
- Revise trainer's version of formative evaluation tool. Create separate tools to capture evaluation of content as separate from evaluation of delivery.
		Start: FY09-10 (Deadline: December 2009)

DESCRIPTION OF POSSIBLE OPTIONS	RESOURCES NEEDED	PROJECTED TIMEFRAME
D. Other: Casework Supervision module of Supervisor Core *Purpose:* To ensure that the evaluation of the training fits with the course revisions. *Recommendation:* Revise embedded evaluation to coincide with revisions that will be made to this module.		Start: FY09-10
E. Other: SDM/Critical Thinking *Purpose:* To assess trainee skill in key aspects of safety assessment. *Recommendation:* Pilot embedded evaluation for SDM/Critical Thinking training which incorporates the SDM Hotline tools.		Start: FY09-10

II. Vertical Projects: Developing Level 6: Transfer or Level 7: Outcomes projects

DESCRIPTION OF POSSIBLE OPTIONS	RESOURCES NEEDED	PROJECTED TIMEFRAME
A. A new transfer of learning on Safety & Risk Assessment following SDM/Critical Thinking Curriculum. *Purpose:* To better assess a key practice skill across a number of evaluation levels. *Recommendations:* - Explore a feasibility plan for TOL piece (including issues of fidelity of implementation). This includes determining available resources (e.g., connections with others such as CRC, Todd Franke, etc.) - Macro Eval Team would make decision about whether to proceed with TOL development for this curriculum only after the pilot of the curriculum and a feasibility plan are completed.		Start: FY11-12

DESCRIPTION OF POSSIBLE OPTIONS	RESOURCES NEEDED	PROJECTED TIMEFRAME
B. An outcomes-related evaluation related to _____ (knowledge/skill area) *Purpose:* To better understand the impact of training on client outcomes. *Recommendations:* Develop an evaluation process, likely tied to new or existing program evaluation efforts. The goal would be to develop a common method of structuring evaluations that could be used on a variety of topics in the future. • Macro Eval Team to develop and complete a feasibility/ resource study, which would be linked to the statewide research agenda (with training variables that are relevant). Possibly add training evaluation component under someone else's research. A decision to move forward on implementation would be made only after this study.		Start: FY09-10

III. Mining Existing Data

DESCRIPTION OF POSSIBLE OPTIONS	RESOURCES NEEDED	PROJECTED TIMEFRAME
A1. Child Welfare Worker Core Demographic data (ESL, race, and gender) *Purpose:* To understand the effects of ESL, race and gender on performance in order to then explore a variety of possible remedies to reduce/eliminate performance differences based on these variables. *Specific issues to be addressed as part of this effort:* - ESL (Define what we mean by ESL: Spoken language? Written language?) - Need to include differential functioning by ESL status in item analysis for all knowledge tests		CWW: FY09-10

DESCRIPTION OF POSSIBLE OPTIONS	RESOURCES NEEDED	PROJECTED TIMEFRAME
A2. Supervisor Core Demographic data *Purpose:* To better understand who supervisors are and how their demographic characteristics are related to their learning. *Recommendation:* - Proceed with development of demographic profiles of Supervisor Core data (similar to what is done for line workers). - Demographic analysis of supervisor evaluation data.		Spvr Core Demographic Profiles: FY09-10 Spvr Core Demographic Analysis: FY11-12
B. Training evaluation Title IV-E data *Purpose:* To better understand the impact of the Title IV-E educational experience on readiness for practice when compared to other new trainees. *Specific questions to be answered:* - In which content areas are Title IV-E trainees improving? - In which content areas are Title IV-E trainees NOT improving? (and then give feedback to specific Title IV-E programs) - In which content areas do Title IV-E trainees come in with more knowledge? - Compare Title IV-E status and time on the job. - Compare results with findings from Title IV-E workforce/career path data. - What content can we put in e-learning so when we do classroom it's more application training?		Start: FY10-11, and across IV-E programs
C. Problematic items *Purpose:* To identify and rectify items that are performing poorly. To have available a bank of items that perform well from which to develop test forms. *Recommendation:* Proceed with item analysis (differential functioning).		Start: FY09-10

IV. New Dimensions: Developing evaluation beyond the existing Framework

DESCRIPTION OF POSSIBLE OPTIONS	RESOURCES NEEDED	PROJECTED TIMEFRAME
A. An attitudes/values evaluation *Purpose:* To understand if and how training impacts values and attitudes. *Recommendation:* Conduct an evaluation of attitudes/values. ▪ Option: Conduct an evaluation of attitudes/values using CMI1. Might be able to combine with evaluation of a neglect scenario if CMI is changed this way (Toby case). Change ethnicity of child and administer to different trainees (possibly using a design that controls for trainee differences that also might relate to their decision making) to see if there are significant differences in how neglect is assessed based on ethnicity. ▪ Option: Conduct an evaluation of attitudes/values using CMI2. (Note: A regional study is about to begin using the Child Forensic Attitude Scale developed by Mark Everson of the University of North Carolina.)		CMI1 Neglect scenario Start: FY09-10 CFAS Start: FY09-10
B. Trainer evaluation (and impact of trainer development on performance) *Purpose:* To better understand trainer differences in training standardized curricula in order to target trainer support efforts. *Recommendation:* Conduct trainer evaluations. - Separate out trainer information from content evaluation on the formative evaluation tools. - Use Common Core test data to examine trainer related differences in performance on specific items and/or competency areas. (Would require large Ns and likely utilize multiple years of data). Could obtain evaluative information via classroom observation and use of other kinds of formative evaluation tools. *Note: It is possible to use retrospective data as long as trainers in a given regional training academy consent to this; otherwise rely on prospective test data.* - Develop model of trainer evaluation with several dimensions.		Start: FY09-10

DESCRIPTION OF POSSIBLE OPTIONS	RESOURCES NEEDED	PROJECTED TIMEFRAME
C. Stereotype threat *Purpose:* To better understand whether ESL, gender and race performance differences can be understood in terms of stereotype threat and, if needed, to make changes in testing and training to address this. *Recommendation: Conduct an evaluation of stereotype threat.* ▪ Pilot within one or across two RTAs/IUC where comparison can occur re: demographic form is administered to training classes before tests vs. demographic form is administered to training classes after tests. ▪ Southern RTA volunteered to pilot and compare cohorts since they do pre-service training and the data collection period is shorter.		Start: FY09-10 (or beyond)
D. Other training modalities: E-Learning *Purpose:* To better understand the relative strengths of e-learning and classroom learning. *Recommendations:* *1*. Include e-learning in the strategic plan for Macro Eval, and coordinate relevant e-learning development efforts with the E-Learning Subcommittee of STEC. 2. Proceed with evaluation of e-learning at the level of formative evaluation – which includes evaluation of both content and delivery platform. Decisions: - No statewide evaluation of e-learning at the knowledge level. - Do not compare e-learning modality with classroom training on same content, since we are not proceeding with a statewide evaluation of e-learning at the knowledge level at this time. - Topic chosen for an e-learning course would likely be a Phase II Common Core content area (those topics which are currently standardized at the level of competencies and learning objectives). Take existing course and put on e-learning format (e.g., AB490 from LA) - Define 'product': it is both curric content & affiliated software used to deliver the product.		Start: FY10-11

DESCRIPTION OF POSSIBLE OPTIONS	RESOURCES NEEDED	PROJECTED TIMEFRAME
E. Other: Quality Assurance for Training *Purpose:* To better understand the impact of training on client outcomes. *Recommendation:* Conduct quality assurance efforts at the statewide and regional levels. ▪ Tailored observer tool would be used. ▪ Small group of same people would observe trainings for one Phase 1 up for revision and one Phase 2 course up for LO revision per year throughout the 5 regions (small group would be composed of statewide reps, regional reps, and county staff dev and liaison personnel; could be a subcommittee of Macro Eval Team; process would be somewhat similar to what was done at the beginning of Core, where the same 7 people observed all of the pilots of the Common Core courses so that data can be gathered in a consistent manner). ▪ Two levels of QA: 1st level – Small group goes to all regions to observe trainings while using observer tool (observer tool refined as part of this process). 2nd level – Members of the small group then go to the regions and train personnel (e.g., RTCs in Central RTA) on use of the tool, so that ongoing regional QA occurs. ▪ This process should fit hand-in-hand with the trainer evaluation process noted above; the intent is not to be punitive toward trainers - instead any observations of trainers would contribute to a professional growth plan.	1. Staff time for: ▪ Level 1 (statewide) QA efforts ▪ Level 2 (within regions) QA efforts 2. Travel reimbursements for: ▪ Statewide reps ▪ RTA/IUC reps ▪ County staff dev reps Per BLJ, travel reimbursements are available for QA process; however, it would also depend on scope of efforts.	Start: FY10-11

V. Other Efforts

DESCRIPTION OF POSSIBLE OPTIONS	RESOURCES NEEDED	PROJECTED TIMEFRAME
Marketing and education *Purpose:* To ensure that key stakeholders understand the benefits and findings of training evaluation and to promote their input. *Recommendation: YES, create brief, targeted, marketing pieces* - Determine targeted marketing groups: trainers, managers, CWDA and Children's Committee, university partners and faculty, Title IV-E PCs - Write brief (1-2 page) pieces using non-technical evaluation language (Mary G. volunteered to assist with this) for targeted marketing groups. - Develop targeted FAQ documents based on questions stakeholders would ask		Start: FY09-10

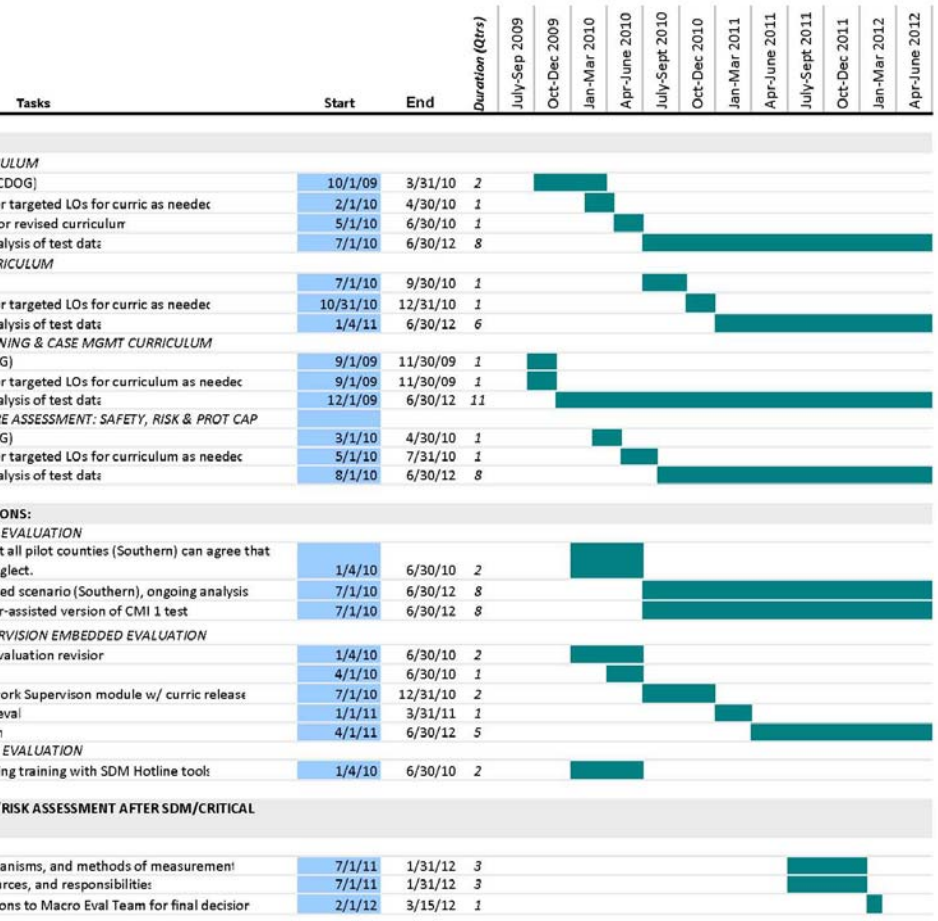
Appendix P: Strategic Plan for Child Welfare Training Evaluation 2009-2012, Gantt Chart

FY 2009-2012 Training Evaluation Strategic Plan (Gantt chart version, FINAL)
California Child Welfare Training - Macro Evaluation Subcommittee of STEC

WBS		Tasks	Start	End	Duration (Qtrs)
1		**FORMATIVE EVALUATIONS:**			
1.1		*OBSERVER VERSION OF FORMATIVE EVALUATION*			
	1.1.1	Develop and obtain input on separate tools for eval of content and deliver	10/1/09	12/31/09	*1*
	1.1.2	Pilot use of revised observer evaluation tools	12/1/09	12/31/09	*1*
	1.1.3	Roll out revised observer evaluation tools	1/4/10	6/30/12	*10*
1.2		*TRAINER VERSION OF FORMATIVE EVALUATION*			
	1.2.1	Develop and obtain input on separate tools for eval of content and deliver	10/1/09	12/31/09	*1*
	1.2.2	Pilot use of revised trainer evaluation tools	12/1/09	12/31/09	*1*
	1.2.3	Roll out revised trainer evaluation tools	1/4/10	6/30/12	*10*
1.3		*E-LEARNING FORMATIVE EVALUATION*			
	1.3.1	Develop and obtain input on tool that captures content and platform	10/1/09	12/31/09	*1*
	1.3.2	Pilot use of draft e-learning formative evaluation tools	1/4/10	6/30/10	*2*
	1.3.3	Roll out e-learning formative eval tool (in conjunction with e-learning curric	7/1/10	6/30/12	*8*
1.4		Identify content that can be put in e-learning to help refine classroom training	7/1/10	6/30/12	*8*
2		**DEMOGRAPHIC INFORMATION:**			
2.1		*CHILD WELFARE WORKER CORE*			
	2.1.1	Define ESL term for further analysis	3/1/10	5/31/10	*1*
	2.1.2	Modify demographic form based on refinement of ESL term	6/1/10	7/1/10	*1*
2.2		*SUPERVISOR CORE*			
	2.2.1	Provide profiles of supervisor core demographic data (ongoing)	10/1/09	6/30/12	*11*
	2.2.2	Provide demographic analysis of supe core test data (depends on tests rec'd	7/1/11	6/30/12	*4*
2.3		*IV-E DATA*			
	2.3.1	Provide annual demographic analysis of IV-E trainee test data	10/1/10	10/31/10	*1*
	2.3.1.1	In which content areas are IV-E trainees improving?	10/1/10	10/31/10	*1*
	2.3.1.2	In which content areas are IV-E trainees NOT improving?	10/1/10	10/31/10	*1*
	2.3.1.3	In which content areas do IV-E trainees come in with more knowledge?	10/1/10	10/31/10	*1*
	2.3.1.4	Compare IV-E status and time on the job.	10/1/10	10/31/10	*1*
	2.3.1.5	Compare IV-E training eval data with Title IV-E workforce/career path data	10/1/10	10/31/10	*1*
2.4		*PROBLEMATIC TEST ITEMS (every September)*			
	2.4.1	Analyze test items for differential functioning based on trainee demographic	8/1/10	9/15/10	*1*
2.5		*STEREOTYPE THREAT*			
	2.5.1	Southern RTA pilot demographic form administered pre- and post-training	10/1/09	9/30/10	*4*
	2.6.1	Provide analysis of pilot data and recommendations	10/1/10	12/31/10	*1*

2009-2012 Training Evaluation Strategic Plan (Gantt chart version, FINAL)
California Child Welfare Training - Macro Evaluation Subcommittee of STEC

WBS		Tasks	Start	End	Duration (Qtrs)
3		**KNOWLEDGE TESTS:**			
3.1		*PERMANENCY & PLACEMENT CURRICULUM*			
	3.1.1	Weight targeted LOs for curriculum (CDOG)	10/1/09	3/31/10	*2*
	3.1.2	Develop, review, and finalize items for targeted LOs for curric as needec	2/1/10	4/30/10	*1*
	3.1.3	Develop test based on targeted LOs for revised curriculum	5/1/10	6/30/10	*1*
	3.1.4	Roll out revised tests and ongoing analysis of test data	7/1/10	6/30/12	*8*
3.2		*CHILD & YOUTH DEVELOPMENT CURRICULUM*			
	3.2.1	Target LOs for testing (CDOG)	7/1/10	9/30/10	*1*
	3.2.2	Develop, review, and finalize items for targeted LOs for curric as needec	10/31/10	12/31/10	*1*
	3.2.3	Roll out revised tests and ongoing analysis of test data	1/4/11	6/30/12	*6*
3.3		*FAMILY ENGAGEMENT IN CASE PLANNING & CASE MGMT CURRICULUM*			
	3.3.1	Weight targeted LOs for testing (CDOG)	9/1/09	11/30/09	*1*
	3.3.2	Develop, review, and finalize items for targeted LOs for curriculum as needec	9/1/09	11/30/09	*1*
	3.3.3	Roll out revised tests and ongoing analysis of test data	12/1/09	6/30/12	*11*
3.4		*CRITICAL THINKING IN CHILD WELFARE ASSESSMENT: SAFETY, RISK & PROT CAP*			
	3.4.1	Weight targeted LOs for testing (CDOG)	3/1/10	4/30/10	*1*
	3.4.2	Develop, review, and finalize items for targeted LOs for curriculum as needec	5/1/10	7/31/10	*1*
	3.4.3	Roll out revised tests and ongoing analysis of test data	8/1/10	6/30/12	*8*
4		**SKILLS-DURING-TRAINING EVALUATIONS:**			
4.1		*NEGLECT SCENARIO AND EMBEDDED EVALUATION*			
	4.1.1	Revise CMI1 scenario (Toby) such that all pilot counties (Southern) can agree that scenario represents neglect or not neglect.	1/4/10	6/30/10	*2*
	4.1.2	Roll out new test form based on revised scenario (Southern), ongoing analysis	7/1/10	6/30/12	*8*
	4.1.3	Regional pilot (Southern) of computer-assisted version of CMI 1 test	7/1/10	6/30/12	*8*
4.2		*SUPERVISOR CORE, CASEWORK SUPERVISION EMBEDDED EVALUATION*			
	4.2.1	Convene workgroup for embedded evaluation revisior	1/4/10	6/30/10	*2*
	4.2.2	Revise embedded evaluation	4/1/10	6/30/10	*1*
	4.2.3	Pilot embedded evaluation for Casework Supervison module w/ curric release	7/1/10	12/31/10	*2*
	4.2.4	Analyze emb eval data & refine emb eval	1/1/11	3/31/11	*1*
	4.2.4	Roll out revised embedded evaluation	4/1/11	6/30/12	*5*
4.3		*SDM/CRITICAL THINKING EMBEDDED EVALUATION*			
	4.3.1	Pilot emb eval for SDM/Critical Thinking training with SDM Hotline tools	1/4/10	6/30/10	*2*
5		**TRANSFER OF LEARNING OF SAFETY/RISK ASSESSMENT AFTER SDM/CRITICAL THINKING TRAINING:**			
5.1		*FEASIBILITY STUDY*			
	5.1.1	ID data sources, data collection mechanisms, and methods of measurement	7/1/11	1/31/12	*3*
	5.1.2	ID logistics: timeframes, access, resources, and responsibilities	7/1/11	1/31/12	*3*
	5.1.3	Provide summary and recommendations to Macro Eval Team for final decisior	2/1/12	3/15/12	*1*



2009-2012 Training Evaluation Strategic Plan (Gantt chart version, FINAL)

California Child Welfare Training - Macro Evaluation Subcommittee of STEC

WBS	Tasks	Start	End	*Duration (Qtrs)*
6	**OUTCOMES EVALUATION:**			
6.1	*FEASIBILITY STUDY*			
6.1.1	ID possible topics linked to research projects from statewide research agenda	7/1/10	1/31/11	*3*
6.1.2	ID data sources, data collection mechanisms, and methods of measurement	7/1/10	1/31/11	*3*
6.1.3	ID logistics: timeframes, access, resources, and responsibilities	7/1/10	1/31/11	*3*
6.1.4	Provide summary and recommendations to Macro Eval Team for final decision	2/1/11	3/15/11	*1*
7	**OTHER TYPES OF EVALUATIONS:**			
7.1	*ATTITUDES/VALUES EVALUATION: CMI 1*			
7.1.1	Vary race/ethnicity of child in CMI 1 scenarios	12/1/09	12/31/09	*1*
7.1.2	Review variations of CMI 1 scenarios.	1/4/10	3/31/10	*1*
7.1.3	Administer variations of CMI 1 scenarios.	4/1/10	6/30/10	*1*
7.1.4	Analyze and report out on CMI1 attitude/values evaluation data	7/1/10	6/30/12	*8*
7.2	*ATTITUDES/VALUES EVALUATION: CMI 2 (CFAS)*			
7.2.1	Collect CFAS data.	9/1/09	12/31/11	*10*
7.2.2	Enter and clean data.	1/4/10	3/31/12	*9*
7.2.3	Analyze and report out on CFAS data	4/1/12	6/30/12	*1*
7.3	*TRAINER EVALUATION*			
7.3.1	ID trainer related differences in item difficulty re: LOs for trainer development	7/1/09	6/30/12	*12*
7.3.2	Develop & obtain feedback on model of trainer eval with several dimensions	10/1/09	6/30/12	*11*
7.4	*QUALITY ASSURANCE EFFORTS*			
7.4.1	Convene small group of reps from around the state (each region represented	7/1/10	8/31/10	*1*
7.4.2	Small group attends one Phase 1 training (curric up for revision) in each region	9/1/10	3/31/12	*6*
7.4.3	Small group attends one Phase 2 training (LOs up for revision) in each region	9/1/10	3/31/12	*6*
7.4.4	Analyze feedback from statewide reps for Phase 1 observations	7/1/11	8/31/11	*1*
7.4.5	Analyze feedback from statewide reps for Phase 2 observations	7/1/11	8/31/11	*1*
8	**MARKETING AND EDUCATION**			
8.1	*MARKETING PIECES*			
8.1.1	Draft brief (1-2 page) docs using non-technical eval language for targeted grps	10/15/09	11/15/09	*1*
8.1.2	Obtain feedback from Macro Eval Team	11/16/09	12/11/09	*1*
8.1.3	Revise marketing pieces	12/15/09	1/15/10	*1*
8.1.4	Disseminate 1st round of marketing pieces	1/15/10	6/30/12	*10*