

Strategic Planning for Child Welfare Training Evaluation in California

Cindy Parry, *Ph.D.*, Jane Berdie, *M.S.W.*, and Barrett Johnson, *LCSW*

Abstract

This paper provides an overview of California's progress in formulating a strategic plan for statewide training evaluation. It outlines general considerations in strategic planning in a large, decentralized training system such as California, and identifies decisions and progress made to date by the collaborative team charged with formulating the strategic plan.

Background Information

In 2002 the California Department of Social Services (CDSS), the California Social Work Education Center (CalSWEC), the Regional Training Academies (RTAs), the Inter-University Consortium in Los Angeles (IUC), and county representatives formed the Macro Evaluation Team to develop and implement a multiyear, comprehensive plan for training evaluation in child welfare. The primary purpose was to assess the degree to which training across the state is effective in preparing direct service child welfare services staff for their work.

During 2003, the federal Child and Family Services Review (CFSR) was conducted in California, and this impacted this strategic planning effort. Work from the Macro Evaluation Team was included in California's Program Improvement Plan (PIP), and additional work was planned as a result of the CFSR/PIP process. Specifically, the California PIP identified two related tasks/action steps:

- In consultation with CalSWEC, CDSS will develop a common framework for assessing the effectiveness of

training that is aligned with the federal outcomes (CDSS, p.220)

- CalSWEC and the RTAs will utilize the results of the evaluation of the models of mentoring to develop a mentoring component which will be included in the supervisory common core curriculum (CDSS, p.222)

Building a Strategic Plan

At the beginning of the strategic planning process, general considerations for the strategic plan were reviewed and discussed by the Macro Evaluation Team.

General Considerations in Strategic Planning

Purpose

One of the first planning decisions is what purpose(s) training evaluation is intended to serve; what kind of information is needed by stakeholders? For example, does the agency (or agencies) want to use evaluation data to improve the training through feedback about trainee learning and transfer in general, or to demonstrate that individual participants have met minimum competency standards? Is there a need to demonstrate training's role in achieving agency goals or affecting client behavior? These decisions impact resource needs and drive several other decisions, including: the levels of evaluation specified by the plan; recommended confidentiality and reporting procedures; and prioritization and timeframes for the evaluations. It is important before beginning to specify what will be evaluated, at what level and when, to be sure that all stakeholder groups have input to the process and that clarity is achieved regarding training evaluation's purpose and scope.

Use and dissemination of data and results

A second key consideration is how the evaluation data will be reported and used. One must decide whether or not individual results will be shared with supervisors or administrators or only aggregate data will be shared (e.g., average results for all or a group of trainees). Evaluation tools that provide individual level data on mastery of a competency or competencies are developed differently and have different resource implications than tools that are designed to give feedback on course effectiveness. If individual data will have any potential personnel consequences, an additional level of rigor is needed to validate the evaluation measures it is based on, to ensure that they are accurate and fair reflections of

knowledge and performance. It may be desirable in a strategic plan to include some guidelines for the rigor of the evaluation design and validation of the evaluation tools that are tied to the purpose and uses of the data. It may also be desirable to include guidelines pertaining to confidentiality and information about the evaluation that is provided to trainees.

Needed Resources

Training evaluation comes with a set of resource needs such as design and analytic skills, database development capacity, and hardware and software needs. It also typically involves new roles for trainers, MIS, and clerical staff. The costs associated with these resource needs vary with need for infrastructure development, the outcomes you want to measure, complexity of the evaluation design, the number of training participants or courses you want to evaluate, who conducts the training evaluation, and existing resources and capacities. Since resources are always limited, a central feature of any strategic plan is a system for prioritizing evaluation activities. In choosing and prioritizing courses or competencies for evaluation several factors should be considered (Parry and Berdie, 1999). These include:

- *The purpose of the course.* Is the training course's purpose to produce changes in behavior, or in attitudes that might be expected to manifest themselves in changes in behavior? Training aimed at producing work-related outcomes (for example, new technical skills or implementation of new policies) is generally a higher priority for evaluation than training being used as a "perk," or as an organizational rite of passage.
- *Cost of the evaluation.* It is important to know what time, money, and expertise will be required for an evaluation, and what resources are available. For example, routinely collected administrative data may help reduce costs of collecting agency- or client-level outcome data.
- *Centrality of the subject matter to competent job performance.* For example, core training might be viewed as critical, particularly if new workers continue to enter the system with widely varying educational and experiential backgrounds.
- *Consequences of a poorly developed and delivered training.* An additional consideration might be the consequences attached to less than competent trainee

performance. These consequences may be to clients. For example, it might be critical to establish worker competency in the area of risk and safety assessment for children. Consequences may also be financial. For example, Webster and Finney (1990) point out failure to do formative evaluation throughout curriculum development is especially risky in Computer-Based Training, since revisions to content are very costly after the initial programming is done.

- *Number of trainees reached.* Training which involves large numbers of trainees and is offered several times a year generally is a higher priority for evaluation than training that is offered to only a few trainees.
- *Length of course.* Courses may be prioritized for higher level evaluation based on the amount of time available for conducting an evaluation. For example, a half-day course might best be evaluated only with a satisfaction survey. A more time consuming knowledge or skills examination might take too much time from the training day.
- *Legal mandates, public and political pressures.* It may be necessary to show the effectiveness of particular training courses in response to lawsuits, settlement agreements or legislative and public concerns. Evaluations conducted for these reasons typically are a high priority and also require more costly and sophisticated outcome level designs.
- *Agency goals.* Connection of training evaluation to larger evaluation efforts/agency goals, and building administrative support by demonstrating the value of training may require consideration of training evaluation at higher levels. In these cases it will likely be necessary to establish, at a minimum, that trainees' knowledge and skills have increased following training, and it may be necessary to demonstrate impact on agency goals, client outcomes, or community goals.

Priorities and Revision

Finally, timelines for the plan should be discussed and agreed upon. These decisions include how long a time period the plan will cover, timelines for phasing in each component, and a schedule for revisiting and updating the plan in response to emerging priorities.

Strategic Planning in California

Using the general framework outlined above, the Macro Evaluation Team began formulating a strategic plan specific to California.

Purposes

As noted above the overarching purpose of a strategic planning process for training evaluation is to establish a plan for what will be evaluated, at what levels, and when, and how the information will be used. During the initial discussion, however, a number of additional purposes were identified that a strategic plan for training evaluation in California might serve. These purposes have been grouped below under five major themes listed below.

The group envisioned a strategic plan as providing a vehicle for:

- *Increased effectiveness and consistency in the provision of training.* A consistent approach to evaluation of training statewide will increase the ability of all training entities to provide the most effective training, thus improving client outcomes. Standards may also be developed that are consistent with the most effective methods.
- *Increased knowledge for stakeholders.* In California, as in other states, state and county administrators often consider training as the solution to many program problems that it cannot impact. Education for County Directors and others regarding training limitations would help to clarify what training can impact and what it cannot.
- *Increased integration of training evaluation with agency, state and federal priorities.* A plan for statewide training evaluation would offer a structured response to the CFSR/PIP process, and could form a basis for evaluating training for future initiatives and reform efforts.
- *Increased cooperation and resource sharing.* Consistent use of a statewide approach to training evaluation would allow curricula to be developed in similar formats, with evaluation integrated within it. This would allow for increased sharing of curricula and evaluation data.
- *Clarification of roles and responsibilities.* A strategic plan for evaluation of training will help to clarify the roles of the counties, CalSWEC, the universities, CDSS, and the RTAs/IUC.

Concerns

In the discussion the group also voiced a number of concerns to be kept in mind throughout the development of the plan. They identified that the strategic plan must be realistic and that implementation should be incremental. The plan should provide useful feedback to all major stakeholders in a timely way so that momentum is not lost. The group also felt that the plan should be integrated with other statewide initiatives and priorities for child welfare, but remain responsive to changing priorities and needs throughout the system. There was concern that the plan should balance a need for consistency statewide with county needs and strive to remain responsive to both. Finally, there was discussion of the role that training can be expected to play in bringing about organizational and systems change and concern that the plan clearly reflects that training is not a panacea for solving all organizational problems.

Goals

The Macro Evaluation Team identified six long term goals for the strategic plan that follow from the purposes for the plan envisioned in the earlier discussion:

Goal One: Establish a “chain of evidence”

The group agreed that in laying out what competencies/courses would be evaluated and at what levels, the plan should be designed to build a *chain of evidence* to support training’s effectiveness. The chain of evidence would support linking training to outcomes such as retention and could provide a foundation for calculating return on investment (ROI). There was agreement that all participants (RTAs/IUC, counties and CalSWEC) would commit to utilizing the approach of building the chain of evidence, however, there was also recognition that not all counties or RTAs would necessarily have the need or the resources to pursue the chain of evidence all the way to evaluations of client outcomes or ROI.

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 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).

Goal Two: Identify common outcomes

A related goal is to identify common outcomes for training participants that the members of the group would agree were high priorities for evaluation for all stakeholders, and to focus efforts to build the chain of evidence around these priorities. Since course offerings vary throughout the state, it is likely that these common priorities will be identified in the form of competency statements to be addressed rather than specific courses or modules.

Goal Three: Integrate federal review findings with training evaluation

In setting priorities for evaluation and determining at what level to evaluate, the group felt that it is also important that the strategic plan be designed to enable the state to respond to findings of the federal reviews now being conducted (Child and Family Service Reviews, IV-E and AFCARS).

Goal Four: Provide an evaluation of the current infrastructure

Another major goal for the planning process is to identify existing infrastructure that would support evaluation of training as well as gaps that would need to be filled. An example of an infrastructure consideration is whether or not a county or RTA has a computerized system for tracking who attends what training. The group felt that it would be necessary to survey the various localities providing training to determine resource availability.

Goal Five: Provide a basis for obtaining statistics regarding delivery and utilization of training

The group discussed the dilemma faced by the state in responding to outside enquiries and reviews when no common

database exits for tracking numbers of people trained and on what topics. Participants agreed on the need for a goal of implementing a common structure for collecting data for basic statistics (one that can be easily utilized). However, a number of practical problems and barriers were mentioned that will need to be explored in implementing this goal.

**Goal Six: Develop a system for sharing resources—
“evaluation central”**

It became clear in the discussion that counties and RTAs vary with respect to their infrastructure needs and resources, and that there is no need for training evaluation to constantly “re-invent the wheel.” Participants suggested that the strategic planning process could provide a vehicle for identifying ways in which resources could be shared and technical assistance provided. For example, an item bank of multiple choice test questions would be developed and validated for a specific group of core competencies that the counties and RTAs could pull from to develop tests tailored to their curricula. Another example would be if one entity had expertise to help the others in highly specialized areas, e.g., involving design, test construction or data analysis.

Progress in California’s Strategic Plan

Multiple Levels of Evaluation

The team began by developing a common understanding of levels of evaluation, and design issues at each level of evaluation. The levels of evaluation currently being conducted by RTAs and counties were assessed and discussed, and work was begun on several levels simultaneously.

Knowledge

The team decided to support the development of a knowledge level evaluation. Consultant Anita Barbee and her colleagues developed a bank of items which were reviewed extensively by RTAs, several counties, and CalSWEC. The team decided to focus on items that are relevant to five areas: child maltreatment, risk assessment, human development, case planning/case management, and placement. The team and consultants have made many revisions over the course of this year. Additional items were contributed by the IUC and the states of North Carolina and Colorado and these too are being reviewed and revised. A secure process is being developed by which counties and RTAs will be able to access the item bank and select relevant items, allowing

them to develop unique tests for their training modules. Tests will be scored by CalSWEC. County and RTA-specific data will be returned only to the those counties/RTAs. Only statewide data will be available to CDSS. Software for item banking and item analysis will be purchased by CalSWEC.

In order to support item development, CalSWEC and its consultants conducted technical assistance workshops on item writing, test construction and item analysis. These workshops have covered topics such as how to select items to maximize content validity, how many items to include for a reliable test, and item writing do's and don'ts to help participants both write and review potential item bank items.

Transfer/Mentoring Evaluation

A second priority area identified by the team was to provide assistance to RTAs and counties related to specific needs and interests. It was recognized that not all participants would have the need or resources to move to more advanced levels of evaluation of training at the same time. The group desired to support evaluation at higher levels without requiring it of all participants.

Evaluation of mentoring programs for new workers was one area of interest. Representatives from the mentoring programs in the Northern and Central training Academies have participated in developing an evaluation design in two phases. In the first phase, begun September 2003, change in worker feelings of efficacy with relation to case skills, supervisors' ratings of worker's case skills, and worker-supervisor relationships are being measured for Core training cohorts exposed to the mentoring models and the comparison group. Research in Texas (Baumann, Kern, McFadden and Law, 1997b) has pointed to the importance of feeling that one has strong case skills in mitigating job stress, turnover, and burnout, and also to the importance of a supportive relationship with a supervisor in developing these feeling of efficacy. Another purpose of the first phase is to gather qualitative and quantitative feedback on the characteristics of an effective mentoring program. The Southern and Central Academies have agreed to collect comparison group data for phase 1. In phase 2, the Academies are interested in collecting data on the mentors' effect on the transfer of learning in a specific skill area, such as case planning. At this time, the feasibility of a common skill based evaluation across the Northern and Central Academies is being explored.

Building on Multiple Levels of Evaluation over Time

The team has begun development of a framework for a strategic plan for evaluation. One approach that has been discussed is to build the plan on the five core priority areas identified in the initial meetings. The group decided to begin with case planning/case management. An example has been developed to illustrate what the evaluation of case planning might look like at varied levels of evaluation and what resources and agreements are needed to implement these evaluations. This example will serve as a working document for the team to respond to, as well as a guide for consideration of the other priority areas.

Implications of Training Evaluation

Curriculum Implications

For evaluation of training to produce *statewide* findings, there need to be some significantly broad areas of commonality among the training entities. In California, the RTAs, the IUC, and some large county training programs have all developed content independently. CalSWEC created a comprehensive curriculum based on all of the curricula in use in the state, but this is more of a resource than a widely used curriculum. There are also no common competencies for line worker core curriculum that are used statewide.

Evaluation of knowledge, however, requires some degree of common content—in order to ensure that the same information is being used for a given topic. The PIP calls for a common core curriculum, which should aid in the evaluation efforts by specifying standard competencies for all core training, and by specifying the content (i.e. the information presented) in the areas of training that will be evaluated using the item bank.

Bank of Knowledge Items

Development of a large knowledge item bank provides an opportunity for each training entity (RTA or county) to select items that are relevant to their content. In order to produce statewide results as well as to assess item reliability, knowledge items will need to be incorporated into tests by at least several training entities. Items that are most commonly selected would help to determine what content was most commonly presented. Another method would be to develop common expectations of content and focus item development on these content specifics.

Skill/Application of Knowledge Evaluation

Evaluation of skill across training entities (statewide or some subset of statewide) requires not only commonality of content but also a common training methods. For skill to be evaluated (whether in the classroom or in the field), trainees must have sufficient opportunity to 1) observe and analyze the skill being used properly, and 2) practice and receive feedback. Moreover, because skills evaluation is most meaningfully embedded in an exercise, practice experience need to be structured in such a way as to provide a common basis of experience leading to the evaluation component. Skill training takes time and evaluation should focus only on the most critical skills in order to use classroom time for skill training judiciously. Again, as part of the PIP, California has chosen one area of skill, identifying child injury/maltreatment, as the area where skill-level evaluation will be standardized.

Culture of Training/Cultural Implications

Training evaluation is currently not part of many child welfare training cultures, except at the level of providing feedback about satisfaction and opinion. Trainers generally do not expect organized classroom time to include embedded evaluation, much less to play major roles in conducting evaluation activities. Most are not trained for this. Trainees do not expect to participate in evaluation and many worry about personnel implications. Agencies may not be ready to participate in evaluation, to support the time it takes for workers' supervisors to be involved in evaluation, or to use evaluative data usefully. Agencies may expect to receive worker-specific evaluation data that the evaluation is not designed to produce. The culture of evaluation needs to developed. A culture shift is needed for training evaluation to be successful.

Resource Issues

Subject Matter Expert (SME) Time

Training evaluation requires significant input of SMEs (i.e., trainers, curriculum writers, caseworkers, and supervisors). An example is the extent of review needed on knowledge items. The same will be needed for skill evaluation designs, e.g., development of “anchors” for scaling—for example, what constitutes an acceptable level of skill demonstration in writing a goal and related objectives in a case plan.

Evaluation Consultation

To date, consultation has been used to support the Team's work in the following areas: facilitate team decision making;

educate training providers about evaluation; develop a knowledge item bank; develop the process for data analysis, management and reporting; select item banking and item analyses software; and develop a mentoring evaluation design. If decisions are made to move forward with evaluation efforts at the skill, transfer or other higher levels, there will be continued need for expertise in evaluation design, development of reliable and valid measures, and analysis of outcome data.

Software

Software needs to be selected/purchased for 1) item banking and 2) analysis. Item banking software costs begin at about \$1000 and can go as high as \$6000. Analysis software such as SPSS starts at about \$600 for a one user, academic price, but is likely to be available already in university settings.

Training for Trainers

Trainers need to be trained and supported in understanding and participating in training evaluation.

Database Management and Data Analyses

CalSWEC staff time is needed to maintain databases and monitor data collection, as well as for scoring and analysis of training data. While consultants will help with initial stages of these processes, it makes sense for the ongoing work to be conducted by staff.

Conclusion

Strategic planning for any project within California's county-administered child welfare system is a complex undertaking. This is particularly true in the case of training evaluation, since multiple entities conduct training and have developed different curricula to meet regional needs. The CFSR/PIP process also complicates the task, by changing the context of the planning efforts while they were under development. To deal with these issues, CDSS, CalSWEC and the RTAs/IUC are planning for a system that will maximize flexibility for counties and university-based training entities, while still providing meaningful feedback on the effectiveness of training at several levels. Work that was begun as part of the Macro Evaluation Team was also integrated into the PIP.

Several strategies were employed to address these complexities. Decisions about the purposes and goals of training were established early on in the process, in order to focus the work. In addition, work was begun on several different levels at once, so

that benefits could be seen throughout the planning process. For example, several entities began work on evaluation of transfer, even while the bank of items to test knowledge acquisition was under development. Potential resources needs were also outlined as the planning moved forward.

References

- Parry, C & Berdie, J. (1999). *Training Evaluation in the Human Services*. Washington, D.C.: American Public Human Services Association.
- Baumann, D., Kern, H., McFadden, T., & Law, J. R. (1997b). Individual and organizational factors in burnout and turnover: A decision-making ecology approach. In D. Bauman & H. Kern (Eds.), *Worker improvements to the Structured Decision and Outcome Model*. Austin, TX: Texas Department of Protective and Regulatory Services
- Webster, S. and Finney, J. (1990). Formative Evaluation and interactive training development. *Bulletin of the American Society for Information Science*, August-September, 19-21.