

ASSESSING TRAINING OUTCOMES: FINDINGS FROM THE NATIONAL EVALUATION OF CHILD WELFARE TRAINING GRANTS

Mary Elizabeth Collins

Boston University School of Social Work

Acknowledgement: This study was funded by a federal training grant to the author from the U.S. DHHS, Administration for Children, Youth and Families, Children's Bureau. Comments regarding this manuscript should be addressed to the author at 264 Bay State Road, Boston, MA 02215, by email, mcollins@bu.edu, or by phone (617)-353-4612.

Introduction

Training of child welfare workers is a core technology aimed at improving services for children and families, and therefore, their outcomes. Despite the extensive federal, state and local resources devoted to child welfare training, this field has received surprisingly little scholarly attention. The fields of child welfare practice and training delivery have both become highly aware of the need for better evaluations to demonstrate outcomes as well as to improve practice. Yet, numerous challenges remain.

Reports of training evaluations in the literature are typically about specific training interventions (Freeman & Morris, 1999; Saunders & Anderson, 2000) and generally have small sample sizes, although a few larger evaluations (with sample sizes greater than 100) are reported (Leung & Cheung, 1998; Jones, Packard, & Nahrstedt, 2002; Mills & Yoshihama, 2002). Challenges in the measurement of skill development and transfer of skills to the work setting are frequently noted difficulties in the training evaluation literature. Common findings from studies suggest that transfer of training is enhanced when trainees are given the opportunity to practice the skill and receive feedback about their performance within the work setting (Baldwin & Ford, 1988; Salas & Cannon-Bowers, 2001). Thus, good supervision is central to performing learned skills in the work setting. The findings of these reviews are applicable to child welfare training; the few studies that have specifically addressed the transfer of training within child welfare systems have generated similar results (Curry, Caplan, & Knuppel, 1994; Gregoire, Propp, & Poertner, 1998; Antle & Barbee, 2003).

In 2003, the Children's Bureau of the U.S. Department of Health and Human Services funded Boston University School of Social Work to conduct the National Evaluation of Child Welfare Training Grants. This national training evaluation project was the first known attempt to evaluate child welfare training beyond individual projects. It focused on a cluster of nine projects with the same basic focus—training workers to help youth with the transition from care—but that occurred in different contexts and with different training methodologies.

Other articles from the National Evaluation project have reported on the context of the training within child welfare service delivery (Collins, Amodeo, & Clay, 2007a), the conceptualization of training (Collins, Amodeo, & Clay, 2008), training delivery (Amodeo, Collins, & Clay, 2008), and youth involvement in training projects (Clay, Amodeo, & Collins, 2008). The focus of this paper is on the evaluations conducted by these training projects and the outcome results reported.

Method

Because training projects were completed or near completion at the time the National Evaluation was funded, a retrospective design was utilized to assess the training projects. The evaluation utilized a multiple case study design; a multiple case study is defined as “empirical inquiry that uses more than a single case in investigating a contemporary phenomenon within its real-life context when the boundaries between phenomenon and context are not clearly evident and in which multiple sources of evidence are used (Yin, 1984).” After conducting a pilot study at one site in the Summer of 2004, the field period for data collection was from August 2004—December 2004. Table 1 provides information on the data collection activities at each site. These activities included interviews, observations, and review of curricula, videos, reports, and other project products. Sites are denoted in this table and the text by letters (A, B, C, etc.) rather than by the grantee name or location. Additional details of the evaluation methodology can be found in Collins, Amodeo, and Clay (2007b).

Findings

Process of Conducting Evaluation

Although an evaluation was required by the original Request for Proposals, no parameters for the evaluation design or procedures for conducting the evaluation were specified. Consequently, there was substantial variation in projects’ processes of evaluation. In three projects, the evaluation was conducted by an evaluation unit within the same organization as the grantee; in four projects, an outside evaluation consultant was hired for at least part of the project. In two cases, the evaluation was conducted by project staff. This was the Principal Investigator (F) in one case and a combination of staff in the other (A).

The level of integration of the evaluation into the overall project design and implementation was variable; in some cases it was well-integrated, in others partially integrated, and in others it appeared quite separate. At Site C, the integration seemed to be particularly strong. Multiple respondents reported an extensive process of collaboration between the curriculum development/training team and the evaluation unit. This process was reportedly helpful in clarifying the curriculum/training design and ensuring the evaluation matched the project’s goals. Conversely, other sites specifically noted that the evaluation was separated from the design of the training project (B, D) and this was believed to be detrimental to useful evaluation.

Table 1: Data Collection Activities/Products Reviewed

	Interviews	Training Observation	Product Reviewed (N)
A	Project director (1), Trainer (1), Representative, grantee org. (1)	None	Grant proposal (1), Final report (1), Interim report (6), Curriculum (12), Media (1)
B	Project director (1), Trainer (2), Curriculum developer (2), State collaborator (1), Evaluator (1)	7 hrs.	Grant proposal (1), Final report (1), Interim Reports (6), Curriculum (1), Media (1)
C	Project director (3), Trainer (1), Curriculum Developer (2), State Collaborator (2), Evaluator (2), Youth (3), Representative, grantee org. (5)	3 hrs.	Grant proposal (1), Final report (1), Interim report (6), Curriculum (3), Media (3)
D	Project director (1), Trainer (1), Evaluator (1), Youth (2), Representative, grantee org. (3)	None	Grant proposal (1), Final report (1), Interim report (6), Curriculum (1), Media (1)
E	Project director (1), Trainer (3), State collaborator (2), Evaluator (1), Representative, private child welfare agency (1)	6 hrs.	Grant proposal (1), Final report (1), Interim report (6), Curriculum (1), Media (1)
F	Project director (2), State collaborator (3), Representative, child welfare agency (2)	None	Grant proposal (1), Final report (1), Curriculum (1), Media (5)
G	Project director (1), Trainer (2), Curriculum developer (2), State collaborator (2), Evaluator (1), Representative, grantee org. (1)	50 min.	Grant proposal (1), Final report (1), Interim report (6), Curriculum (6), Media (4)
H	Project director (1), Trainer (1), State collaborator (1), Youth (1), Representative, grantee org. (2)	None	Grant proposal (1), Final report (1), Interim report (6), Curriculum (3), Media (4)
I	Project director (1), Trainer (2), State collaborator (2), Youth (2), Representative, grantee org. (2)	5.5	Grant proposal (1), Final report (1), Interim report (6), Curriculum (1), Media (1)

There was a wide range of approaches to conducting the evaluation and in some sites there was a more concerted attempt at evaluation than in others. Across all projects, the evaluation that was implemented was different, sometimes highly so, from that which was planned as described in the proposal. Two sites (C, E) seemed to have improved their evaluation strategy over the course of the project. For example, at Site E the evaluation that was conducted initially consisted of a pre-test/post-test developed by project staff that was primarily focused on satisfaction. There were, however, several problems with the original instrument designed by

project staff without evaluator input: it was too long and had too many questions, the language of the questions was problematic, and project staff was unsure what to do with the data. It became clear to the project staff that the instrument did not work, so they eventually hired an outside evaluator to conduct the evaluation. This was considered to be far more successful.

Four sites (D, G, H, I) explicitly stated in their proposal a plan to utilize Kirkpatrick's (1994) four level evaluation model (participant satisfaction, participant knowledge, participant skill, and impact on worker effectiveness and/or client progress). It is well known that evaluation at the higher levels is most challenging to conduct. Sites were ambitious to plan comprehensive evaluations, but all had difficulty implementing evaluation at the higher levels. Each of these four sites attempted to collect follow-up data in an effort to measure worker effectiveness and/or client progress but there were several challenges to follow-up assessments. These included dependency on the public agency to contact trainees at follow-up, to secure a control group, and to collect data regarding performance or impact. Other reasons given for scaling back the scope of the original evaluation plan included: technology problems, project staff transitions, curriculum design changes (i.e., as the curriculum/training were developed this led to changes in the evaluation), and time/budget constraints.

In part because these were all demonstration projects, substantial evaluation activity was devoted to formative evaluation in order to gather feedback needed to improve and finalize the curriculum and training development. In general, less effort was given to outcome evaluation. The efforts that projects gave to outcome evaluation and the methods used were widely disparate. Even within projects, different methods were used at different phases of the project and for different training modules or audiences.

Most projects relied on a standard pre-test/post-test design to evaluate knowledge, attitude, or perceived impact on skill. There was substantial variation on measures. Examples of some of the measures used are described below.

- Two sites utilized retrospective measures (C, E). These types of measures collect data after the training but ask the respondent to reflect on what they knew before they participated in the training (i.e., "now that you know [specific information], think back to what you did not know before training...").
- Site F developed a competency measure that was used at pre-test and post-test and focused on skill development. Consistent with the content of the training the measure was designed to assess four areas of competency: partnering, strengths-based responding, building youth autonomy, and support building. Three measures were used, each containing case vignettes to assess application of knowledge and attitude.
- At Site I, participants rated five statements on a 5-point scale: I am satisfied with the training; I will incorporate some of what I learned into my work; I

will do a better job because of this training; My agency will support me in using knowledge/skills from this training; Adolescents and their families will benefit from my taking this course. Measurement occurred at post-test and three months follow-up.

The most comprehensive evaluation conducted was at Site C, described in detail below:

Case Highlight: Site C

Part of the explanation for the evaluation sophistication at Site C was because the grantee organization was a Center which had an existing evaluation unit as part of the Center infrastructure. An initial step to the evaluation was the development of an evaluation logic model. This was a combined process (over six months) of involving the evaluation team with the training project team. Following this the evaluators conducted a process evaluation for eight months that involved the completion of reaction forms by training participants and the observation of training by the evaluators. To evaluate outcomes the evaluation utilized a retrospective pre-test/post-test and a three-month follow-up.

The evaluation team produced a detailed evaluation report that was separate from the project final report. The report provided information on three levels of outcomes: reactions; changes in knowledge, skill, attitudes, and intentions; and change in practices. The evaluation report included an overview of the evaluation, the logic model, the instruments, and results. In addition to a reaction questionnaire (used to guide curriculum/training development), two outcome questionnaires were designed. One was administered immediately after the training and one was administered three months after the training. The questionnaires were designed to measure: 1) immediate changes in participants' knowledge, skills, attitudes, and intentions; and 2) intermediate-term outcomes that involved changes in practices or immediate gains. Because the project developed three different curricula for different audiences, with its own distinct outcomes, three unique sets of outcome instruments were created. Each set contained a retrospective pre-test questionnaire and a follow-up questionnaire. This project also made a concerted effort to conduct a three-month follow up. Still, however, the response rate was disappointing. Details about the outcome evidence are provided in the next section.

The evaluation was not without challenges; the evaluation team identified the following as key challenges to conducting the evaluation: amount of time spent developing the logic model, limited effort to monitor training fidelity, the state budget crisis led to decreased participation in training, and high turnover among training participants which affected follow-up.

Outcomes Evidence

This section provides a summary of outcome data (quantitative information on knowledge, attitude, or skill) provided by the projects. The final reports of the projects were highly variable in the amount of outcome data provided. Two projects

(B, I) provided virtually no outcome data. At the other end of the continuum, Site C had a full separate evaluation report in addition to the project final report.

The Site C project evaluation report provides information on three levels of outcomes: 1) reactions (trainees' reactions to training content, quality of instruction, amount of support they believe they have to implement what is presented in the training, and their overall satisfaction with the training experience); 2) changes in knowledge, attitudes, skills, and intentions; 3) changes in practices. Evaluation data were collected for all three targets of training (supervisor, caseworker, and youth professionals). Small sample sizes, particularly at follow-up, were problematic for the supervisor and youth professional training. The sample for caseworkers was sufficiently large ($n=110$) but only 36 percent returned the follow-up survey. For these reasons, evaluation data should be cautiously interpreted.

The evaluators summarized the outcome results as follows. First, the training produced immediate gains in knowledge, skills, and intentions across all three levels of training (e.g., caseworkers, supervisors, youth workers). Second, most immediate gains were maintained, and there was evidence to indicate that caseworkers changed their work-related practices (e.g., in terms of practice, all trainees who responded at follow-up reported using the tools and techniques they learned in the training). Efficacy scores regarding the use of independent living tools and understanding of issues of trauma and loss significantly increased from pre-test to post-test and were maintained at follow-up.

Two other sites (F, H) provided fairly detailed outcome information. These data are summarized below:

- At Site F, the focus of the evaluation was on a series of skills: partnering, strengths-based responding, building youth autonomy, and support building. The final report provides detailed data on these domains for 26 participants. The data suggested that the training was able to produce changes in participants. The strongest area of change was youth-worker partnership and the interpersonal relationship. Uneven gains appear in the area of youth empowerment. There was also some improvement in how the participants worked with support systems.
- At Site H, the final report provided pre-test/post-test and six month follow-up evaluation data. Data were provided on pre-test/post-test (knowledge test and competency assessment) for four cohorts of trainees and for a group of supervisors/managers and one Training of Trainers. Significant gains were recorded for most scores. Data on six month follow-up was provided for 25 participants and indicated positive results.

Evaluation data were provided by four other projects, but the reports were briefer.

- Site E used a retrospective pre-test with trainees. Data were provided for 195 social workers completing the retrospective pre-test. The evaluation tool contained 16 outcome-related questions ranging from a measure of social workers' increasing understanding of "the impact that demonstrated caring, concern, and attention for social workers has on transition age foster youth development" to "the resources available to parenting foster youth." The final report lists the quantitative measure on each of these objectives that suggests change from pre-test to post-test. The conclusion of the evaluation report cautions, however, that there is no evidence that the project changed practices for participating trainees.
- At Site G, the final report provided data on post-test (trainee satisfaction, perceptions of the training's relevance to their work, perceptions of their own learning for each competency area) for ten cohorts of trainees on the core module, two cohorts of trainees on the culture module, and three cohorts of trainees on the mental health and substance abuse module. The evaluation found that many trainees gained substantial knowledge and skills to effectively work with adolescent populations. For example, in the mental health and substance abuse module, 70 percent of trainees got at least 70 percent of items correct in the first two sessions. The grantee conducted a follow-up phone interview for one of the modules—the culture module—to measure trainee's levels of transfer of training. Among 21 trainees who completed the follow-up interview, 53 percent reported that they had not used the learned ethnographic interviewing and 32 percent stated that they had not used the concepts of youth culture subtypes in their work. In contrast, 15 percent reported that they used learned information on cultural groups.
- At Site D, data from a follow-up survey (24% response rate) showed a positive response to the training including a high percentage of respondents reporting (a) use of the skills and tools learned in the training in their work with youth; (b) agreeing that the advocacy material helped them in their work, (c) agreeing that the networking material helped them in their work, and (d) strongly agreeing that the "adolescent development" material helped them in their work.
- At Site A, very limited evaluation data was provided in the final report. For example, evaluation data comparing the pre-test and post-tests demonstrated an overall 65 percent increase in "knowing how to complete a culturally competent life skills assessment" and an overall 72 percent increase in "knowing how to create a culturally congruent plan of transition."

Two sites (B, I) reported no data in their final reports. Efforts at Site I, to collect 3 month follow-up data were unsuccessful. The intent was for trainees to complete

mailed postcards asking follow-up questions. However, project staff did not follow up with the trainers to see whether these instruments were administered. At Site B, measurement of attitudes and knowledge were attempted. But project staff reported that due to the late point at which the evaluators were involved, the technical difficulties involved in the embedded evaluation, and the fact that the curriculum was delivered by some trainers inexperienced with the technology, no usable evaluation data was produced.

Outcomes: Perceptions

Although projects provided limited documented evidence about impact, project personnel and collaborators perceived the projects to be successful in many ways, most prominently in achieving attitude change regarding adolescents. This feedback was reported at all sites. Below is a sampling of the data collected from interviews with project staff and collaborators describing their perceptions of the impact:

- Workers felt a greater sense of self-efficacy in working with high risk youth, a greater appreciation for youth strengths, a greater awareness of the issue of readiness for change, a greater understanding of concepts presented, and a greater sense of how to approach these youth.
- Workers had increased awareness of resources available through the Chafee Independent Living program.
- Trainees stated that they tried something learned in the training and it worked, or that through the training they started thinking of a new direction to try in their work with youth.
- Trainees' comments on the training included statements like "powerful"; "hit home"; "received a blessing"; "I learned something new"; "I made a connection."
- There was a lot of testimony at the training sessions about the impact of the work—this was described as an "emotional shaking up"; "you reminded me that my job is about human beings."
- The training format created shifts in attitudes as the youth development philosophy was implemented; trainees' observations of youth and adults partnering to provide training was an effective way of creating attitudinal change.
- There was a philosophical shift in worker views of youth; involvement in the project was an "extremely positive experience" and the project was "absolutely fascinating."
- Feedback sheets included comments like the "best training ever attended"; training gave fresh ideas and perspectives.

- A respondent from a large private agency commented that although she can't identify the impact on youth with evidence, she "can't imagine how it hasn't helped the youth" as there is "such a direct correlation between a worker's perspective on youth work and youth outcome."
- The training increased youth credibility due to youth involvement in the training and curriculum development; there was a positive response from trainees for including the voices of youth.
- Many workers and community providers took away a more positive view of youth and will be more likely to treat them as resources.

System Effects

Some projects (B, C, E, I) suggested that there was an impact on the systems, primarily child welfare, which will result in better outcomes for youth in the long run. It was primarily suggested that during the time the training was being delivered there were other youth-focused and Independent Living—related efforts in the states and/or counties. It was the combination of activities that may have had an impact on systems that raise attention to the needs of adolescents in child welfare. Illustrative of this perspective was a comment at Site E that in combination with Chafee grants and Independent Living money, the curriculum "continued the ball rolling" to serve youth better. Also, collegial relationships that developed in the course of the project continued. The assumption is all this activity is bound to have an effect on youth in the long term.

Discussion

Reviewing the data regarding the process of conducting evaluation and the outcome evidence reported leads to a number of conclusions and recommendations regarding training evaluation in federally-funded child welfare projects.

All sites were required to conduct an evaluation of their training projects but there was wide variation in how projects attempted this and most conducted more limited evaluation than originally planned. Sites recognized the importance of evaluation but also the substantial difficulties in designing and implementing evaluations, and in the analysis of evaluation data. Review of the processes of conducted evaluations identified the importance of a designated evaluator. When the role is simply one part of that of the Principal Investigator or Project Director it tends to get less attention. The most successfully conducted evaluation was that of Site C in which a full evaluation unit was engaged. Also important was the integration of evaluation with curriculum design and training delivery functions at an early stage in the process. This led to better evaluation.

Projects appeared overly ambitious regarding evaluation design at the proposal stage. The reason for this is uncertain. State agency involvement in evaluation is needed regarding control groups, follow-up, and potential collection of some types

of outcome data (job performance, supervisor evaluation, client outcome). Perhaps projects assumed assistance from the state agency with these tasks only to find reluctance when they attempted to carry them out. Consequently, securing cooperation for the evaluation design should be negotiated at the proposal phase.

The quantitative evaluative data reported by projects was limited. In part, projects appeared to invest more effort in formative evaluation, so that they might improve the training project, rather than outcome evaluation. Given that these were a series of demonstration projects, this seems like a reasonable use of limited evaluation resources. Yet, what lessons can be learned from these projects' efforts to conduct outcome evaluation? Three issues seem paramount.

First, there are the measurement challenges. Few standardized measures exist in child welfare training evaluation and measures typically need to be specific to the type of training being offered. Projects, therefore, tend to design their own. The field needs to invest more effort in developing measurement tools that can be utilized by training evaluators, and to make them available for use. In this way, training evaluations can build on each other and develop a field of knowledge.

Second, the cooperation of state agencies is needed for any type of sophisticated evaluation design, particularly those involving access to control groups, ability to contact workers after training, and to gather certain types of measurement data (e.g., job performance, or observation of skill in the work setting). Projects often planned for such methodological sophistication, but needed to conduct more limited evaluation when difficulties securing agency cooperation arose.

Third, even for these fairly small, intense, and federally-funded projects the evaluation challenges were significant. This suggests even further difficulties are in store for the evaluation of larger-scale efforts, for example, core training, that are so central to child welfare practice. Even the most well-designed evaluation study with full cooperation will have additional challenges when aimed at training a broader workforce. This is especially true for the measurement of skills and competencies which often require in-depth observation rather than paper-and-pencil self-report. Engaging supervisors in the training enterprise, therefore, is essential. Supervisors in the workplace offer the best potential for assessing skills of the trained workforce over the long term.

There was some evidence that the cluster achieved some outcomes over and above those achieved by individual projects. The cluster helped bring more attention to the issues of youth transitioning from foster care and their special needs. Undoubtedly, this led to further efforts such as the supervisory training projects reported in this volume.

Training interventions often serve multiple purposes. Federally-funded projects such as participating in the National Evaluation of Child Welfare Training Grants may potentially have substantial, unmeasured, long term benefits in addition to an impact on individual trainees' knowledge, attitudes, and skills. Some of the long term benefits suggested by respondents included: enhanced collaborations that influence program and practice development; institutionalization of training in the

agency setting; or advancements in knowledge development for the field regarding better approaches to training. Although these types of benefits are generally not included in the evaluation of individual training projects they are critically important and need further attention.

Conclusion

Training evaluation efforts for individual projects are typically small in scale. The National Evaluation project, although methodologically limited by its retrospective design, was able to provide a larger scale evaluation than is usually possible. Consequently, findings can provide a richer spotlight to training evaluation efforts. Elevation of attention to training evaluation in child welfare is needed, given that training is a core component of child welfare interventions and it receives significant resources. The data reported here and their discussion aim to foster increased scholarly dialogue about training evaluation.

References

- Amodeo, M., Collins, M.E., & Clay, C.M. (2008). *Best practices in curriculum design and training delivery: Examples and gaps in federally-funded independent living training projects*. Unpublished manuscript available from the author.
- Antle, B. & Barbee, A.P. (2003). *Training transfer: Variables that predict and maximize transfer*. National Resource Center on Child Welfare Training and Evaluation: University of Louisville.
- Baldwin, T.T., & Ford, J.K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology, 41*, 63-105.
- Clay, C.M., Amodeo, M., & Collins, M.E. (2008). *Involving youth in training: Lessons for consumer involvement in social services interventions*. Unpublished manuscript available from the author.
- Collins, M.E., Amodeo, M., & Clay, C.M. (2007a). Training as a factor in policy implementation: Lessons from a national evaluation of child welfare training. *Children and Youth Services Review, 29*, 1487-1502.
- Collins, M.E., Amodeo, M., & Clay, C.M. (2007b). Evaluation of Independent Living training grantees: *Case study final report*. Boston: Boston University School of Social Work.
- Collins, M.E., Amodeo, M., & Clay, C. (2008). Planning and evaluating child welfare training projects: Working toward a comprehensive conceptual model. *Child Welfare 85*(5), 69-86.

- Curry, D., Caplan, P., & Knuppel, J. (1994). Transfer of training and adult learning (TOTAL). *Journal of Continuing Social Work Education, 6*(1), 8-14.
- Freeman, K.A., & Morris, T.L. (1999). Investigative interviewing with children: Evaluation of the effectiveness of a training program for child protective service workers. *Child Abuse and Neglect, 23, 7*, 701-713.
- Gregoire, T.K., Propp, J., & Poertner, J. (1998). The supervisor's role in the transfer of training. *Administration in Social Work, 22*(1), 1-18.
- Jones, L., Packard, T., & Nahrstedt, K. (2002). Evaluation of a training curriculum for inter-agency collaboration. *Journal of Community Practice, 10, 3*, 23-39.
- Kirkpatrick, D. (1994). *Evaluating training programs: The four levels*. San Francisco: Berrett-Koehler.
- Leung, P., & Cheung, K.M. (1998). The impact of child protective service training: A longitudinal study of workers' job performance, knowledge, and attitudes. *Research on Social Work Practice, 8*(6), 668-684.
- Mills, L.G., & Yoshihama, M. (2002). Training children's services workers in domestic violence assessment and intervention: Research findings and implications for practice. *Children and Youth Services Review, 24*(8), 561-581.
- Salas, E., & Cannon-Bowers, J.A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52*, 471-499.
- Saunders, D.G., & Anderson, D. (2000). Evaluation of a domestic violence training for child protection workers and supervisors: Initial results. *Children and Youth Services Review, 22*(5), 373-395.
- Yin, R. K. (2003). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.