

The role of training in the evaluation of public programs

Jeanette Treiber, Robin Kipke, Travis Satterlund
and Diana Cassady

Introduction

Nearly all private, government and non-governmental organizations that receive government funding to run social or health promotion programs in the United States are required to conduct program evaluations and to report findings to the funding agency. Reports are usually due at the end of a funding cycle and they may or may not have an influence on the continuation of program funding. The final evaluation report (FER), as the end-of-funding-cycle report is often called, generally relates the intervention and evaluation results of the funding period and has a dual purpose. It is considered an element of accountability (Compton *et al.*, 2001) and should give the program and its stakeholders direction for the future (Patton, 1997). All too often though, this is not the case. Evaluators have voiced myriad concerns about the many issues related to reports and their usage (Hoefler, 2000). In their study of a random sample of American Evaluation Association members, Torres *et al.* (1997) found that evaluators are generally discontent about reporting and about the fact that their reports are often misused or not used at all. The authors (p. 110) report that the three main factors evaluators cited for impeding success in communicating and reporting evaluation findings were 'insufficient time available to devote to communicating/reporting (53 percent)', clients being unclear about their communication and reporting needs (47 percent) and clients and audiences being unresponsive to communication and reporting efforts (40 percent). These facts are especially disconcerting if we consider that many local programs, such as the local tobacco control programs in California, operate on multi-year policy objectives that often build on previous accomplishments and lessons learned. Evaluation reports could be a valuable instrument for moving projects forward if stakeholders and project staff would make good use of evaluation findings (Rossi *et al.*, 2004). We may be facing a vicious cycle whereby clients or programs do not value reports very highly,

□ Jeanette Treiber, Project Director at the Center for Evaluation and Research, University of California, Davis, USA. Email: jtreiber@ucdavis.edu. Robin Kipke, Evaluation Associate at the Center for Evaluation and Research, University of California, Davis, USA. Email: rakipke@ucdavis.edu. Travis Satterlund, Evaluation Associate at the Center for Evaluation and Research, University of California, Davis, USA. Email: tdsatter@ucdavis.edu. Diana Cassady, Associate Professor, Public Health Sciences, in University of California, Davis, USA. Email: dlcassady@ucdavis.edu

Acknowledgement: Funding for the research of this paper was provided by the California Department of Public Health Tobacco Control Program.

which leads evaluators to paying less attention to the task of writing them well, which in turn has an effect on report quality and makes them less user friendly.

The Tobacco Control Evaluation Center (TCEC) (2006) at the University of California at Davis developed scoring measures for final report writing for over 100 local tobacco control projects in California but found 2007 reports lacking in quality. In 2010, it conducted a training campaign in the hope that the projects themselves, the funding government agency and TCEC may make better use of the reports. TCEC wanted to know if local program directors and evaluators were willing to participate in voluntary report writing training, whether FER writing could be improved through training and if the mode of training (1-day face to face or 1-h webinar) made any difference to improvement. The response to the training call was overwhelming, and comparing scores from 2007 and 2010, participating agencies made statistically significant improvements but non-participants did not. Results relating to the mode of training were inconclusive. The pre- and post-score comparison proved to be a valuable measuring tool, and the 1-day face-to-face training was a useful training mode.

Background

California's tobacco control programs are funded through a state tobacco tax. The state disburses funds to approximately 100 projects consisting of county health departments, a few metropolitan areas and community-based organizations for projects predominantly covering second-hand smoke protection through local policies (outdoor areas, multi-unit housing, events, etc.). The state requires that all projects develop 3-year work plans, including process and outcome evaluations. It also funds TCEC, a statewide organization that provides one-on-one technical assistance, training and resources in order to strengthen local projects' evaluation capacity (Treiber *et al.*, 2011). TCEC also reviews and scores the FERs; those projects are required to submit at the end of each 3-year funding cycle. An examination of the FERs submitted in 2007 revealed that many were unsatisfactory overall because they did not conform to standard reporting procedures. The mean score of all submitted FERs was 66.5 percent, and their usefulness for subsequent campaigning was limited. For instance, summary data were often reported without drawing conclusions; evidence was missing that process data had been used to inform the intervention; and recommendations for future direction of the program were often missing. In a nutshell, it seemed that FERs were completed and submitted only to meet a requirement and not to enhance projects' work. Moreover, TCEC also uses the FERs as data sources for summary reports, highlighting statewide accomplishments, challenges and barriers on specific objectives such as multi-unit housing outdoor area policies, etc. As the FERs differed greatly in data quality, analysis and comprehensiveness, it became difficult to aggregate data without further research. The need to make reports conform to standard data analysis and format requirements became another motivation for investing in a capacity building campaign before the next round of FERs was due.

In 2006, TCEC published 'Tell Your Story', containing 'Guidelines for Preparing Complete, High Quality Final Evaluation Report[s]'. Each state-funded tobacco control project in California was advised to use this guide. It contains how-to instructions as well as a sample report, a checklist and the scoring sheet used by the TCEC to evaluate these reports. The guidelines conform to generally accepted methods of writing evaluation reports for health promotion and disease prevention programs (Bamberger, 2006; Moskowitz, 1989; Preskill & Russ-Eft, 2005); scores are given for title page, abstract, project description, evaluation methods, evaluation results, conclusions and recommendations. Reports are not rated according to program achievement but according to whether or not there is evidence that implementation and evaluation activities were carried out as planned, with an emphasis on soundness of evaluation design, sampling decisions, data collection instruments and procedures, data analysis, presentation of findings and conclusions drawn.

Methods

Training activities

TCEC conducted a face-to-face 1-day training in seven locations throughout California. All project directors and evaluators were invited to attend an optional regional training meeting, and travel reimbursement was offered to those travelling more than 100 miles from a training site. Registration was set up on the TCEC website, and training dates and registration information were announced on a statewide listserv, via e-mail to project directors and evaluators, and on TCEC's Facebook page. For each of the trainings, a few participants were contacted in advance with the request to speak about their own experiences and/or anticipated challenges with writing the FER. The 1-day training consisted of a series of presentations on the Purpose of the FER, Anatomy of a FER, Writing the Results Section and Ways of Disseminating FER Findings.

Most time was spent on interactive exercises, which included group work on scoring a section from a FER, and making decisions on what information from an actual evaluation activity to include in a FER. The day ended with an office hour that allowed participants to ask specific questions. Participants received a binder with training materials that included the slides from presentations, exercise handouts and a number of additional resources such as sample FERs and instructions on how to develop a press release from evaluation findings, etc.

TCEC offered additional training events and information before FERs were due in order to complement the regional training meetings. A 1-h webinar had the same content as the face-to-face training but did not include any interactive exercises. An edition of the TCEC newsletter 'Process and Outcome' was entirely dedicated to writing and improving FERs. Moreover, TCEC posted sample evaluation reports that had received high marks in the 2007 round on its website.

Measures and analysis

To determine whether local program directors and evaluators were willing to participate in an optional training meeting or webinar on report writing, TCEC tracked the participation of agencies in the training events and conducted satisfaction surveys at the end of each 1-day training meeting. A satisfaction survey was not performed for the webinar. The unit of analysis for participation was both individual and agency, whereas the unit of analysis for the satisfaction survey was individual participants. To assess whether FER scores improved as a result of participation in the training events, we compared FER scores from 2007 and 2010. TCEC staff scored FERs in 2007 using a written protocol to promote consistency in scoring.

Scoring in 2010, after the evaluation capacity building activities were delivered, used the same written scoring protocol as in 2007. An outside contractor not involved in the training activities scored approximately 60 percent of the FERs, which consisted of agencies that received training and those that did not. The remaining FERs were scored by staff who delivered the training. In order to test for bias, we compared mean scores. The mean scores of the scorer/trainers were similar to that of the outside scorer.

Analysis of changes in FER scores was conducted for all agencies to assess a trend and for the 56 agencies that submitted FERs during both funding periods. These 56 agencies are similar in that they are all local health departments that receive non-competitive renewals for each funding cycle. The FER score used for these analyses was an average of scores from two to four reports submitted by each agency. Most agencies worked on more than one program objective and were required to submit a report for each objective. Paired *t*-tests were calculated to determine improvement.

Finally, we were unable to compare the effects of the in-person meeting and webinar because only one agency attended the webinar only. Instead, we compared the scores from agencies that attended the in-person training, in-person plus webinar and no training.

Results

Participation in training events

A total of 124 participants from 97 agencies took part in the training events. This number was unexpected and exceeded the number of agencies funded in that cycle (91). Some funded agencies did not participate, but new agencies that had received funding for the following 3 years participated, as well as the California Department of Public Health. Other agencies that attended were subcontracting agencies of California Health Departments.

TCEC had offered the webinar in addition to the regional face-to-face training because it anticipated that some agencies might not have been able to attend a full-day event. However, when participation in the webinar and face-to-face training was analysed by agencies, only one agency attended the webinar only. All other (20) agencies at the webinar were also represented at the regional training.

Satisfaction with training

On the training satisfaction survey, participants were asked to rate components of the training on a four-point Likert scale. Of the 124 participants, 96 filled out the evaluation form. Results of the training satisfaction survey show that the training was very well received. Most importantly, 71 (74.7 percent) strongly agreed that the training would help them write a high-quality FER, and most others agreed. Seventy (72.9 percent) strongly agreed, most others agreed that the presentations were informative and 68 (70.8 percent) agreed that the interactive exercises were helpful. Eighty-three (87.4 percent) strongly agreed and most others agreed that the handouts would be a useful reference material. In the comment sections of the survey, participants also mentioned that they appreciated the training as a networking opportunity. The survey results suggest that there was a good balance of presentations, interactive exercises and supplemental materials.

The best of the training exercises received was a hands-on role-playing exercise that put participants in the position of FER scorers. With the scoring sheet in hand, small groups were given 30 min to score a section of the report. The group had to discuss the section and reach a consensus on the score. Afterwards, each group reported on the rationale for providing the score they did. When all section scores were totalled, the total was consistently several points lower than how TCEC had scored in the original report. In follow-up discussions, participants noted that they had a better sense of the requirements for writing the report and an appreciation for the difficulties in deciding on borderline cases (e.g. 'If results are there but in the wrong section, is that a point deduction?' 'If the elements of the analysis are there but the analysis was erroneous, does that make a 1 or 2 point deduction?'). This was an opportunity for the trainers to explain the decisions the scoring team had reached on such questions.

Assessing changes in evaluation report writing

The 56 health departments work on multiple objectives at the same time, whereas community organizations work on one objective only during a funding cycle. The average score of all Health Departments' and community organizations' FERs in 2007 ($N = 181$) was 21.29 out of 32, or 66.5 percent. The average score of all health departments and community organizations' FERs in 2010 ($N = 179$) improved to a score of 24, or 75 percent.

FER scores for local health departments improved over time regardless of participation in training. Out of a maximum possible score of 32, the 56 local health department FERs received an average score of 20.3 in 2007 and 23.6 in 2010. This is a statistically significant improvement ($p = 001$).

Analysis of scores by intervention method

In terms of score comparison by intervention, the average score of health departments that were represented at the 1-day training ($N = 34$) was 25.05 (out of a total possible 32)

and those who attended both ($N = 20$), 22.63. Those who attended no training and no webinar averaged 20.4. Agencies that attended the training improved their scores at statistically significant levels, whereas non-attending agencies did not (Table 1).

Discussion

Improving evaluation report writing on a large scale, as is the case with California tobacco control projects that involve more than 100 agencies, is a challenging task. TCEC attempted to improve the evaluation reporting skills of these organizations through two training methods: a 1-day training of project directors and evaluators, and a webinar. The most astonishing result was that project directors and evaluators willingly participated in the evaluation capacity building events even though for most participants continued funding was not dependent on FER scores. Score comparisons show that overall scores improved significantly ($p = 0.01$) from 2007 to 2010. When looking more closely at the score improvement among those attending the training events and those who did not, statistically significant improvement was only found for those attending a training, whereas those not attending a training also improved slightly, but not at a statistically significant rate. When scores of those attending both training events were analysed, their improvement was not statistically significant, but those who attended the face-to-face training only did. It may be that those attending both events had felt the greatest need for training and were not as skilled to begin with than those attending the 1-day event only. As just one agency attended only the webinar, the sample size was too small to draw conclusions on the webinar and was therefore not analysed.

Using a score comparison to measure capacity building offers more reliable results than self-reporting. Judging from the results of the training satisfaction survey, almost all participants felt very strongly about improved capacity to write their FERs and one may have expected a more dramatic improvement. This indicates that training satisfaction surveys with self-reported knowledge gains may be very limited in measuring actual capacity building. The training exercise that put participants in the role of scorer was the most valued by participants, which confirms Trevisan's (2004) and Patton's (2002) suggestions that hands-on evaluation training is most useful for evaluation capacity building. However, this study was only able to confirm this through self-reported anecdotal evidence from training participants. A more systematic study of training methods would be necessary to correlate training methods with capacity gain.

There are some limitations to this study. Even though evaluation reports were scored by following a strict protocol, scorer bias may have played a role in the improvement of scores. However, we feel confident about the results of our study because an outside scorer was used to score the majority of the reports and because mean scores among TCEC scorers and the outside scorer were similar. Moreover, the outside scorer had no knowledge of which agencies had participated in the training. Another limitation of this study is the limited information we have on the actual training elements that may have caused improved scores for participants. We also do not know whether the special FER newsletter edition or the FER-related materials on the website had an influence on outcomes. Further research is needed to identify the training modes that have the greatest impact on report writing improvement.

Table 1: Comparison of 2007 and 2010 final evaluation report scores by intervention

Intervention in 2010	Number of agencies	2007 scores pre-intervention mean	2010 scores post-intervention mean	t-Test significance
None	14	19.8	20.4	$p = 0.15$
Training only	34	21.04	25.05	$p < 0.05$
Training and webinar	20	19.9	22.63	$p = 0.35$

Conclusion

Many of the reports in the 2007 round did not conform to the standard set by required guidelines. One of the consequences was that TCEC had difficulties using the reports as sources to create statewide summary reports on common objectives that local projects had worked on. Moreover, TCEC doubted the internal usefulness of the reports for the individual projects. For these reasons, TCEC set out to conduct a training campaign before the next round of reports were due. The results are encouraging in that they confirm that statistically significant improvement can be achieved via training. Our first research question, 'Are local program directors and evaluators willing to participate in report writing training?' can be answered in the affirmative. Even though this training was voluntary, it attracted not only a majority of funded agencies but also subcontractors. Some of the agencies took full advantage of both training opportunities. The second question, 'Can FER writing of local programs be improved through training?' is also affirmed. Training can significantly improve evaluation report writing results. For the third question, 'Does the training mode (1-day face to face or 2-hour webinar) make a difference in improvement?', our results are limited by the fact that only one agency participated in the webinar only. It was surprising to see that a number of agencies opted to participate in both the face-to-face training and the webinar. Equally surprising was the result that those who participated in both events did not significantly improve, but those who attended the face-to-face training only did, and overall results for those attending trainings (training only or webinar plus training) also improved significantly. We are cautious in drawing conclusions on the webinar training because participation of only 21 agencies is low. However, we do recommend a 1-day training that includes hands-on scoring practice of sample reports as a useful way to improve report writing. Our second overall objective, improving the ease of creating summary reports from projects' FERs, has only partially been met. A summary report was produced from 2010 FERs on one common objective that included 39 agencies, and the author of the report still had many reservations about the quality of the FERs even though a slight improvement to previous years was seen.

This study shows that training has potential effects in improving the evaluation of public programs but reminds us that improving reporting in large scale projects with multiple implementers comes in small steps. It is therefore all the more important to continue developing training methods and materials for improving the writing of evaluation reports.

References

- Bamberger, M. (2006), *Real World Evaluation. Working under Budget, Time, Data, and Political Constraints* (Thousand Oaks, CA: Sage).
- Compton, D., Baizerman, M., Preskill, H., Rieker, P. and Miner, K. (2001), 'Developing evaluation capacity while improving evaluation training in public health: the American Cancer Society's Collaborative Evaluation Fellows Project', *Evaluation and Program Planning*, **24**, 33–40.
- Hoefer, R. (2000), 'Accountability in action? Program evaluation in nonprofit human services agencies', *Nonprofit Management and Leadership*, **11**, 167–77.
- Moskowitz, J. M. (1989), 'Preliminary Guidelines for Reporting Outcome Evaluation Studies of Health Promotion and Disease Prevention Programs', in M. T. Braverman (ed.), *Evaluating Health Promotion Programs: New Directions for Program Evaluation* (San Francisco, CA: Jossey-Bass), pp. 101–12.
- Patton, M. Q. (1997), *Utilization-focused Evaluation*, 3rd edn (Thousand Oaks, CA: Sage Publications).
- Patton, M. Q. (2002), *Qualitative Research Methods*, 3rd edn (Thousand Oaks, CA: Sage Publications).
- Preskill, H. and Russ-Eft, D. (2005), *Building Evaluation Capacity* (Thousand Oaks, CA: Sage Publications).
- Rossi, P. H., Lipsey, M. W. and Freeman, H. E. (2004), *Evaluation: A Systematic Approach* (Thousand Oaks, CA: Sage Publications).
- Tobacco Control Evaluation Center (TCEC) (2006), *Tell Your Story: Guidelines for Preparing a Complete, High Quality Final Evaluation Report* (Sacramento, CA: California Department of Health

- Services/Tobacco Control Section), Department of Human & Community Development, University of California, Davis, Available at <http://www.cdph.ca.gov/programs/tobacco/Documents/CTCPTellYourStory.pdf> (accessed 16 October 2012).
- Torres, R., Peskill, H. and Piontek, M. (1997), 'Communicating and reporting: practices and concerns of internal and external evaluators', *Evaluation Practice*, **18**, 2, 105–25.
- Treiber, J., Cassady, D., Kipke, R. A., Kwon, N. and Satterlund, T. D. (2011), 'Building the evaluation capacity of California's local tobacco control programs', *Health Promotion Practice*, **12**, 6 Suppl 2, 118S–24S.
- Trevisan, M. S. (2004), 'Practical training in evaluation: a review of the literature', *American Journal of Evaluation*, **25**, 255–72.