The Canadian Journal of Program Evaluation Vol. 19 No. 1 Pages 125–142
ISSN 0834-1516

AN EMPIRICAL STUDY OF BUILDING THE EVALUATION CAPACITY OF K–12 SITE-MANAGED PROJECT PERSONNEL

Paul R. Brandon
Terry Ann F. Higa
University of Hawai'i at Manoa
Honolulu, HI

Abstract: This article examines the effects of professional development, including formal workshops and ongoing consultation, on the evaluation capacity of K–12 school faculty and administrators who were conducting evaluations of 17 site-managed projects. Changes in the faculty's and administrators' (a) attitudes toward evaluation, (b) self-confidence as evaluators, and (c) assessments of their capabilities as evaluators were examined. School personnel's attitudes toward evaluation did not improve, but their self-confidence as evaluators and their assessment of their evaluation capabilities both showed improvement. The conclusions buttress the argument that, with training and the assistance of experienced evaluators, school personnel can build their evaluation capacity. A number of limitations in study design and data are noted.

Résumé: Cette étude examine les effets du perfectionnement professionnel — ateliers formels et consultation continue — sur les capacités d'évaluation du personnel d'un établissement scolaire de niveau primaire et secondaire, qui effectuait l'évaluation de 17 projets autogérés. On a examiné (a) les changements dans les attitudes de l'établissement et des administrateurs envers l'évaluation, (b) leur sentiment de compétence en tant qu'évaluateurs, et (c) l'évaluation de leurs capacités en tant qu'évaluateurs. Les attitudes du personnel scolaire envers l'évaluation demeurent inchangées, mais leur sentiment de compétence en tant qu'évaluateurs et leur évaluation de leurs capacités d'évaluation se sont tous deux améliorés. Les conclusions de l'étude corroborent l'argument selon lequel la formation et l'encadrement par des évaluateurs expérimentés peuvent aider le personnel à améliorer ses capacités d'évaluation. Plusieurs limites sont notées dans la conception et les données de l'étude.

Corresponding author: Paul R. Brandon, Curriculum Research & Development Group, University of Hawai'i at Manoa, 1776 University Avenue, UHS2-214, Honolulu, HI, USA 96822; <brandon@hawaii.edu>

K–12 school administrators and faculties in the U.S. and elsewhere are increasingly conducting state- or district-funded site-managed projects (Mohrmann, Wohlstetter, & Associates, 1994). Under broad funding guidelines, schools conducting site-managed projects identify important school needs, receive funding to address the needs, and use the funding to initiate and conduct projects addressing the needs. The school sites manage their own projects under the loose oversight of the funding entity.

A single funding program might support many school projects, making program evaluation problematic. Conducting many individual project evaluations can be expensive. School project personnel can be required to conduct their own evaluations, but often they lack proper levels of evaluation training and experience.

One approach to ensuring that site-managed project evaluations are adequate is to have project personnel evaluate their projects with the assistance of external consultants who (a) provide the project personnel with professional development (PD) in how to conduct evaluations, (b) monitor the progress of the evaluations and provide assistance with technical issues, and (c) write the project evaluation reports or assist project personnel who are writing the reports. The external evaluators train the school administrators and faculty who are managing the projects in how to conduct evaluation tasks, provide consultation as the evaluations are conducted, and ensure that the reporting is fair and accurate by writing the reports or providing considerable assistance to project personnel as they write the reports.

Using this approach is a means of ensuring not only that site-managed projects are evaluated but also that K–12 school personnel's *evaluation capacity* is built. Evaluation capacity is a topic that is receiving increasing attention among evaluation practitioners and researchers (Compton, Baizerman, & Stockdill, 2002; Cousins, Goh, Clark, & Lee, 2003). It is defined as the

> processes and practices that are in place to make a quality evaluation and its uses routine. To the extent to which evaluation capacity is present within the organization, members of the organization develop their knowledge of evaluation logic and methods. They also develop their skills in actually doing evaluation. (Cousins et al., 2003, p. 6)

This article presents quantitative findings on the effects of evaluation PD on K–12 site-managed school-project personnel's perceptions of their capacity to conduct project evaluations. The article differs from much of the research on evaluation capacity building, which has primarily been qualitative, with emphasis on single case studies (Cousins et al., 2003). The findings of the study presented here are limited in that they do not provide direct evidence of the effects of PD on project personnel's practices as project evaluators. However, this is not unusual in the PD literature as a whole. In a large-scale study of PD in U.S. education, Garet, Birman, Porter, Desimone, Herman, and Yoon (1999, p. 3-2) concluded,

> Over the past decade, a large body of literature has emerged on professional development, teacher learning, and teacher change. … Despite the size of the body of literature, however, relatively little systematic research has been conducted on the effects of professional development on improvements in teaching or on improving students' outcomes.

Although this study does not examine the effects of PD on project personnel's evaluation practices, it nevertheless is a first step in examining the effects of PD on school-project personnel's evaluation capacity and is a necessary component of a body of literature on evaluation capacity building.

The purpose of this article is to present the results of a small study of the effects of PD on K–12 school administrators' and faculty's perceptions about their program evaluation capacity. The three research topics include the extent to which the PD

- helped improve school personnel's attitudes toward evaluation,
- helped improve school personnel's confidence in themselves as evaluation practitioners, and
- helped school personnel perceive themselves as more capable at performing major evaluation tasks.

DESCRIPTION OF THE PD PROVIDED FOR THE EVALUATIONS OF SITE-MANAGED PROJECTS

This study examined the effects on site-managed project personnel of PD in evaluation, including brief workshops and long-term fol-

low-up consultation, that a two-person team of professional evaluators provided in the statewide school district in Hawai'i to K–12 administrators and faculty during one school year. The administrators and faculty served a total of 17 projects, each of which was funded by one of two district-wide programs (*N* projects funded in one program = 10; *N* projects in the other program = 7). Except for two instances, each project occurred entirely within a single school. The two exceptions were a multi-site project that served six schools and a multi-site project that served three schools. Of the two programs that provided funding to the projects that we examine here, one funded projects at schools that had high proportions of students who were at risk of educational failure. About $75 per student was provided to these schools to spend in the manner that the schools deemed best addressed student needs. The second program provided funding to schools (or groups of schools) that were competitively awarded grants for innovative projects addressing school needs. The projects were small; funding ranged from about $10,000 to $50,000 for the year, with no extra funding provided for evaluation.

The school district required that the projects funded by the two programs be evaluated but knew that the personnel for the 17 projects by and large were insufficiently knowledgeable or experienced to conduct the studies without assistance. Therefore, it contracted the authors' organization to provide project personnel with PD in how to conduct their evaluations. Our responsibilities were to conduct workshops or meetings at strategic points during the year and to provide follow-up consultation. The purposes of the workshops and meetings were to provide training in how to define and describe project objectives and methods, design evaluations, and select or develop methods for conducting evaluations. Project personnel were responsible for (a) attending PD workshops or, for projects at distant schools, to meet with us for individual PD; (b) defining and describing their projects' objectives and methods; (c) collaborating with us on developing their evaluation questions, designing their evaluations, and identifying or developing data-collection instruments; (d) collecting data; and (e) consulting with us when they had questions throughout the school year. Sign-in sheets at the workshops and meetings showed a total *N* of 59 participating project personnel; for most projects, two to six people were trained or participated in the evaluations and PD (mean = 3.7, st. dev. = 2.6).

We developed two sets of original materials (totaling 58 pages) for the PD, one discussing how to design evaluations and the other ad-

dressing evaluation methods in depth. We distributed the materials to the project personnel at three-hour workshops or in individual consultations. After the initial PD workshops or meetings were conducted, we provided technical assistance in person and by telephone, fax, and e-mail. We also wrote the schools' evaluation reports.

We did not monitor the topics addressed during the follow-up consultation. However, it is reasonable to assume that the pattern of topics was the same as for the previous year, when we provided PD and consultation and did monitor these topics. During that year, we noted that of 32 projects, 72% asked for recommendations about data-collection methods, 56% needed assistance with identifying the focus of their evaluations, and 56% needed recommendations about data analyses. Other requests were for advice about specifying respondent groups (31%), help with identifying the appropriate comparison (e.g., comparison or control group, pre- to post-test growth, and so forth) (22%), reviews of data-collection instruments developed by the project personnel (22%), advice about commercially published instruments (12%), and other technical assistance (18%).

In the initial project workshop, we distributed a questionnaire asking (among other topics) for demographic data on the 41 participating project personnel. These data addressed the respondents' current school positions, their highest degrees, their formal classes in evaluation-related university subjects, their participation in evaluation seminars or workshops, and their participation in evaluations. The results are shown in Table 1. Nearly half of the group were teachers. The group was highly educated, with about 6 out of 10 having master's or doctoral degrees. The group's formal coursework in topics related to program evaluation was primarily at the undergraduate level, however, and the average number of seminars or project evaluations they had participated in was minimal. Together, these demographic statistics show a group whose primary professional focus was on teaching, with some exposure to evaluation or education research methods.

The level of participation of project personnel varied among projects and across the course of the school year. Some personnel who attended the workshops or initial meetings participated in the evaluation studies throughout the entire school year; other project personnel began participating in the evaluations after one or both of the workshops or initial meetings. This occurred in part because project personnel had considerable latitude in how much they par-

ticipated in the PD and contributed to their project evaluations, in part because of a history in the district of school independence in following district decisions, in part because the funding programs were loosely overseen by district administrators, and in part because we had no authority to require participation.

Participation also varied among projects in two other ways. First, some schools were on other islands in the Hawaiian island chain and were less accessible to us. Personnel from projects on the island

Table 1
School Personnel's (*N* = 41) Demographics Collected at First Workshop[a]

Demographic	Pre-project questionnaire respondents (total *N* = 41)	Post-project questionnaire respondents (total *N* = 20)
	Percentage	
School position:		
Teachers, counselors, or librarians	46%	50%
Curriculum coordinators	20%	15%
Administrators	17%	20%
Other positions	17%	15%
Highest degree:		
Bachelors' degree or professional diploma	37%	35%
Masters' degree	49%	55%
Doctorate10%	10%	
No response	4%	0%
University courses in subjects related to evaluation:		
Undergraduate statistics	46%	40%
Graduate statistics	24%	25%
Undergraduate tests and measurement	61%	60%
Graduate tests and measurement	27%	30%
Undergraduate research or evaluation course	29%	35%
Graduate research or evaluation course	39%	40%
Other evaluation professional development or experience during the past five years:	Mean and standard error of the mean	
Number of short, non-university seminars or workshops on program evaluation (e.g., for Title I programs)	*M* = 2.7 St. dev. = 4.2	*M* = 2.2 St. dev. = 2.4
Number of program or project evaluations or school accreditation studies	*M* = 2.8 St. dev. = 6.8	*M* = 1.9 St. dev. = 2.6

[a] An additional 17 personnel completed only post-project questionnaires, but no demographics were collected on the post-project version of the instrument.

of O'ahu, where we are located, participated in a three-hour workshop. Personnel from projects conducted on neighbouring islands did not attend the O'ahu workshops; instead, we met with these personnel at the individual school sites. Consultation by e-mail, fax, and telephone was provided when requested or when we needed to give feedback to the participating schools. It is possible that the effects of the PD might have varied between the project personnel who attended workshops and the project personnel with whom we met in initial PD meetings. Second, two schools had consultants helping with large projects, which were funded in part by one of the two programs discussed here and in part by other programs. School personnel at these schools needed less PD, and their evaluations were conducted in part by their consultants.

METHODS

Research Topic 1

The first research topic addressed the differences between the participants' attitudes toward program evaluation at the beginning and at the end of the school year during which PD was provided. We addressed this topic in large part because previous research has suggested that evaluations of the sort conducted here are more likely to succeed when school personnel participating in the evaluations have positive attitudes toward evaluation (Cousins & Earl, 1995). We also wished to see if we could improve participants' attitudes toward evaluation.

To address this research topic, we developed a questionnaire for collecting data at the beginning and end of the year during which the PD was provided. When developing the instrument, we pilot-tested it with a small group of volunteers, who made suggestions about revisions in wording and format. The instrument was distributed to all school personnel with whom we met during the first stage of the PD, either at introductory workshops or at individual school meetings. It was distributed to project personnel again at the conclusion of the school year, this time by mail via school principals or project coordinators.

Altogether, a total of 58 project personnel completed the pre-project questionnaire, the post-project questionnaire, or both. A total of 41 respondents completed the pre-project questionnaire. Of these 41 people, 20 (49%) also completed the post-project questionnaires. An

additional 17 respondents who did not complete the pre-project questionnaire completed the post-project questionnaire, for a total of 37 post-project respondents. The 21 project personnel who completed the pre-project questionnaire but not the post-project one either dropped out of the project evaluations during the course of the year or failed to respond to the post-project questionnaire. The 17 project personnel who completed the post-project questionnaire but not the pre-project one began participating in the evaluation after the initial meeting.

As shown in Table 1, the demographics for the 20 pre/post respondents were very similar to the demographics for the group of 41 respondents who answered the pre-project questionnaire. The primary exception was that the 20 pre/post respondents had participated in fewer seminars or workshops on evaluation and in fewer project evaluations, program evaluations, or school accreditation studies. Thus, members of the group that responded to the instrument at both the beginning and end of the year, who are the focus of two of the analyses reported here, had somewhat less exposure to ongoing evaluations and probably had less practical evaluation experience before the project than the full group of pre-project respondents. Other than this exception, we can conclude that the 20 respondents were similar to the group who participated in the project evaluations from the beginning. Demographics for the full post-project group are unknown, because demographic data were collected only on the pre-project instrument.

We collected data addressing the first research topic with a five-item, five-point semantic differential questionnaire of project participants' attitudes toward evaluation. Data were analyzed for the 20 project representatives who participated in the projects throughout the school year and completed both the pre-project and post-project questionnaires. The questionnaire asked the respondents how enjoyable, important, and beneficial they thought evaluation was, how favourable their attitudes toward evaluation were, and how eager they were when thinking about doing evaluations. Cronbach's alpha equalled .87 on the pre-project questionnaire and equalled .90 on the post-project questionnaire. Means were calculated for the pre-project questionnaire and for the post-project questionnaire, a *t*-test of differences between pre- and post-means was calculated, and the effect size, calculated as the post-project mean minus the pre-project mean divided by the mean of the two standard deviations, was produced.

Research Topic 2

Data addressing the second research topic (the extent to which the PD helped improve project personnel's confidence in themselves as evaluation practitioners) were collected as pair comparisons in the second section of the instrument that was used to address the first research topic. The pair-comparison method is a venerable scaling method (Guilford, 1954) that currently is less used in social science or education research or evaluation than it was in its heyday. With this method, a set of words or phrases (technical terms, concepts, adjectives, and so forth) relevant to a research or evaluation topic is identified, and each word or phrase is paired with each other. If X is the number of words or phrases, the number of pairs is $(X (X - 1)) / 2$. For the pair-comparison instrument used in this study, we paired the term *program evaluation* with each of six adjectives, including four positive adjectives describing the project participants' perceptions of evaluation (*engaging*, *manageable*, *understandable*, and *worthwhile*) and two negative adjectives (*difficult* and *confusing*). For each pair, the respondents were instructed to write a number giving their perception of the conceptual differences ("distances") between *program evaluation* and each of the six adjectives. Data for the 18 respondents who completed all of the pair-comparison portions of the both the pre-project and the post-project questionnaires were analyzed.

Usually, pair-comparison data are analyzed and reported using multidimensional scaling techniques or other methods. These methods require larger numbers of respondents and of words or phrases than we had in our study. Therefore, our analysis of our pair-comparison data was rudimentary. The purpose of the analysis was to determine the extent to which the project participants held evaluation in a more favourable light at the end of the school year than they did at the beginning. To accomplish this, we calculated (a) the means of the distances of each of the six adjectives from the term *program evaluation* that the project participants indicated in their responses to the pre-project questionnaire and (b) the means of the distances that they indicated in their responses to the post-project questionnaire. We also calculated the means of the differences between the pre-project and the post-project distance scores. To the extent that the results showed decreases in mean distances between *program evaluation* and each of the four positive adjectives from the beginning to the end of the year, the evaluation PD could be said to have had a favourable effect on project personnel's percep-

tions of evaluation, and to the extent that the results showed increases in mean distances between *program evaluation* and each of the four positive adjectives from the beginning to the end of the year, the evaluation PD could be said to have had an unfavourable effect. The reverse was true for the negative adjectives. The data were not normally distributed; therefore, we used non-parametric methods to calculate the statistical significance of the differences between pre-project and post-project distances. We also calculated the effect size for each adjective. Although effect sizes are most appropriate for normally distributed data, we report them here because they help give a sense of the practical significance of the results.

Research Topic 3

The third research topic addressed the extent to which the workshops and consultation helped project personnel perceive themselves as more capable of performing major evaluation tasks. This topic is closely related to the second topic. Data addressing it were collected on a subset of the group of project personnel who provided data addressing the first and second topics. As part of a larger qualitative study (Higa, 2004), we used a retrospective method in interviews after the conclusion of the school year to ask 14 participants from eight schools about their perceptions of the PD they had received during the previous school year. Only respondents from the larger of the two site-managed programs examined in this study were interviewed. The 14 respondents were interviewed because they had participated the most fully in the evaluation PD over the course of the entire school year. On a scale of 1 to 10, with *1* = not at all and *10* = very well, the participants were asked by an independent interviewer to assess their knowledge of how well they could conduct three major evaluation tasks—defining or describing project objectives and methods, writing evaluation plans, and collecting evaluation data—*before* they were trained. They then were interviewed at length about their recollections of the PD in each of the three major evaluation tasks; at the end of the interview, they were asked to indicate on the 1–10 scale how well they could conduct each of the three tasks by the end of the year. In this article, only the quantitative pre/post self-ratings are reported. (The qualitative findings are forthcoming [Higa, 2004].) Means were calculated of the pre-project ratings and of the post-project ratings, and pre/post *t*-tests and effect sizes were calculated.

RESULTS

Research Topic 1: Participating Project Personnel's Attitudes toward Evaluation

The total pre-project attitude-scale results are fairly high (mean = 20.3, st. dev. = 3.9, with a maximum possible total score = 25), and the post-project results are slightly lower (mean = 20.0, st. dev. = 3.6). The difference between the pre- and post-results is not statistically significant (t = .56). The effect size for this decline in attitudes toward evaluation is .08, which is a trivial difference.

We speculated that the high pre-project mean scores might have reflected prior exposure to formal education about evaluation and research, prior PD in non-university seminars on the topics, or previous evaluation experience. To address this speculation, we regressed pre-project attitude-scale total scores of the 20 respondents who completed both the pre- and post-project questionnaires on (a) highest degree (expressed as an ordinal variable, with bachelor's degree = *1*, professional diploma = *2*, master's degree = *3*, and doctoral degree = *4*), (b) the total number of types of formal research or evaluation courses completed, (c) the number of non-university seminars or workshops in evaluation during the past five years, and (d) the number of evaluation or school-accreditation studies. No statistically significant relationships were found; the four independent variables accounted for only 8% of the variance in pre-project attitude scores. These results suggested that prior exposure to evaluation or research did not predispose project personnel to positive attitudes toward evaluation. (Indeed, the variables *highest degree obtained* and *number of university courses* showed slightly negative [-.11 and -.23, respectively] correlations with pre-project total attitude scores.)

Research Topic 2: Participating Project Personnel's Confidence in Themselves as Evaluators

The results of the analyses of the pair-comparison data are given in Table 2. All the effects were in a direction indicating that project personnel's perceptions of program evaluation were more favourable after the project than before. The distances between the term *program evaluation* and the four positive adjectives were less on the post-project questionnaire than on the pre-project questionnaire; the reverse was found for the two negative adjectives. The results

are statistically significant at the .05 level for each of the six adjectives and for the total difference score. The largest differences were for the adjectives *manageable*, *understandable*, and *worthwhile*, with effect sizes of -.52, -.59, and -.61, respectively. Although there are no hard-and-fast rules about interpreting the meaning of effect sizes, effect sizes of these magnitudes are usually considered moderately-sized effects. The effect sizes for *engaging*, *difficult*, and *confusing* were all an absolute value of .35 or less, which are considered small-sized effects.

Research Topic 3: Project Personnel's Perceptions of their Capabilities as Evaluators

The results of the analyses for the third research topic are shown in Table 3. The results of the *t*-tests show that the differences between the 14 interviewed participants' average pre- and post- ratings were statistically significant at the .05 level for each of the three major evaluation tasks. As might be expected, the mean pre-project rating for *defining and describing project objectives and methods* was higher than the ratings for the other two evaluation tasks; it is reasonable to expect that the project personnel were more familiar with describing their projects than they were with other evaluation tasks.

Table 2
Results of the Analysis of Mean Differences in Perceptions of "Distances" Between the Term *Program Evaluation* and Six Adjectives about Evaluation

Adjective about program evaluation	Pre-project perception of distance between *program evaluation* and the adjective			Post-project perception of distance between *program evaluation* and the adjective			Effect size[a]	Difference between pre-and post-perceptions of "distances"[b]	
	Mean	St. dev.	Median	Mean	St. dev.	Median		Mean	Median
Engaging	61.9	39.9	57.5	49.4	31.5	50.0	-.35	25.8	20.0
Manageable	46.1	32.8	50.0	31.4	23.3	27.5	-.52	31.9	30.0
Understandable	47.5	33.0	50.0	31.6	20.9	25.0	-.59	29.8	22.5
Worthwhile	29.4	34.2	15.0	14.7	14.3	10.0	-.61	23.0	15.0
Difficult	43.3	32.9	45.0	51.7	39.0	45.0	.23	-28.3	-20.0
Confusing	55.5	31.3	50.0	65.8	38.5	50.0	.30	-33.0	-30.0

[a] Calculated as (mean post – mean pre)/average of the two standard deviations.
[b] Two nonparametric tests (the signed-rank test and rank-test) each showed that the differences between the pre- and post- perceptions of "distances" were statistically significant for each of the six adjectives.

The change from pre- to post- was small for this task; the effect size was .31. The pre-project rating (3.4) for *collecting evaluation data* was somewhat lower, indicating that the interviewees were somewhat less confident in their capabilities to collect data. The effect size for this task was .59, which can be considered moderate. The pre-project mean (2.6) for *writing evaluation plans* was the lowest of all three tasks, indicating that the project personnel were least confident about this task. The effect size for this variable was .88—a large effect. Notably, the post-project means were all at about the same level, ranging from 4.5 to 4.8, indicating that the project personnel felt similarly confident about all three tasks. Their confidence levels were not high, however: all the post-project ratings were slightly below the halfway point on the 1-to-10 scale.

CONCLUSIONS

With data that go beyond the usual simple retrospective descriptive accounts of evaluation capacity building, we present findings suggesting that PD can have a positive effect on project personnel's opinions about their evaluation capabilities. The study does not, of course, address a full model of the variables affecting evaluation capacity, but it is a step toward that direction and helps broaden the empirical foundation upon which future research can be built. Even though the study is small, the findings on the first two research topics are

Table 3
Results of Analyses of Interviewees' Ratings of Their Capabilities to Conduct Each of Three Major Evaluation Tasks Both Before and After the Evaluation Professional Development (PD)

Evaluation task	*N* interviewees	Perceptions of capabilities before the PD		Perceptions of capabilities after the PD		*t* value[a]	Effect size[b]
		Mean	St. dev.	Mean	St. dev.		
Defining and describing project objectives and methods	13	4.0	2.8	4.8	2.3	3.11	.31
Writing evaluation plans	14	2.6	2.3	4.5	2.0	5.77	.88
Collecting evaluation data	14	3.4	2.3	4.6	2.0	3.77	.59

[a] All results were statistically significant at the .05 level.
[b] Calculated as (mean post – mean pre)/average of the two standard deviations.

strengthened by the fact that they are uniformly statistically significant, despite the low statistical power due to small *N*s. Furthermore, the considerable time lapse between the administration of the pre-project and the post-project questionnaires makes it unlikely that the findings were influenced by a memory effect and suggest that the strength of the workshops, the consultation, and the school personnel's experience evaluating their projects were durable.

The positive attitudes that project personnel had toward evaluation before receiving the PD do not reflect prior PD or experience in evaluation, but they might reflect attitudes typical of school personnel who choose to take part in evaluations and to persist throughout the evaluation studies. It is plausible that these people are likely to have positive attitudes toward evaluation—perhaps too positive to expect improvements after PD during one school year. We speculate that school administrators and faculty choosing to take part in evaluation training and to conduct project evaluations might be unlikely to commit themselves to the task or to remain for the duration unless they have positive attitudes from the beginning.

The findings addressing project personnel's confidence in themselves as evaluators suggest that we can expect PD, including both formal workshops and ongoing consultation, and participation in evaluations to improve project personnel's confidence in themselves as program evaluation practitioners. Furthermore, the findings on improvements in project personnel's perceptions of their capabilities as evaluators suggest that we can expect trainees to report improvements in their capabilities to perform major evaluation tasks.

It should be noted that project personnel did not rate themselves as highly capable. Clearly, more than a single year of PD and participation in project evaluations is necessary if evaluation capacity is to be increased significantly.

It might be enlightening to examine the effects of the PD alone by disentangling the effects of the workshops and consultation from the effects of participating in the project evaluations, but this is impossible in this study. Suffice it to conclude that the findings support the efficacy of the PD in contexts such as the one studied here.

LIMITATIONS AND SUGGESTED FUTURE RESEARCH

The findings reported here broaden our understanding of the effects of PD on improving K–12 school personnel's understanding of evalu-

ation principles and methods and on school personnel's practical application of these methods to evaluating school projects, site-managed or otherwise. The study's conclusions about the perceived effects of PD buttress the argument that school personnel conducting site-managed projects can build confidence in themselves—in the present case, with the assistance of experienced evaluators. These gains in self-confidence enhance the likelihood that evaluation capacity building will improve schools and their projects.

Considerable future research on PD in evaluation and evaluation capacity building should be conducted. The present study examined only some aspects of PD. In addition to examining affective issues, such as those examined here, studies should use direct measures to collect data on changes in project personnel's evaluation knowledge and skills. Research also is necessary on the demonstrated effects of PD on actual evaluation practice. The education research literature on PD of all sorts rarely reports quantitative findings on the ultimate effects of the PD; the evaluation capacity building literature is no different in this regard. The community of researchers studying evaluation capacity should examine the credibility, validity, and usefulness of the findings of project evaluations that are conducted by school personnel after they receive evaluation PD. Long-term studies that follow project personnel for more than a year would be valuable, particularly in instances when these personnel continue to participate in project evaluations year after year. Context issues (Higa, 2004) should also be studied further; such research might help us understand how well school personnel apply the results of PD when conducting evaluations in other settings.

The low percentage of project personnel who responded to both the pre-project and post-project questionnaires limits somewhat the strength of the conclusions about this study's first two research questions. We know from the comparison of demographics between the full set of pre-project respondents and the subset of pre/post respondents that the two groups were quite similar; these results suggest that we can reasonably generalize from the pre/post group to the full pre-project group. However, we do not know how many school personnel who responded to the pre-project questionnaire dropped out of the study, or, if they did, why. Conceivably, some project personnel might not have participated throughout the school year because they were not learning enough or because they were not sufficiently motivated. The most valid conclusion is that the findings generalize to personnel who participate in PD and ensuing

project evaluations throughout the entire PD and evaluation period. Future capacity-building research on changes in attitudes toward, and confidence in, evaluation should track project personnel's participation more closely and collect qualitative data on the participants' experiences and reasons for persisting in or dropping out of the studies.

The timing of the data collection addressing the third research topic might have affected the validity of the results. The respondents were asked for their opinions about the extent to which they could conduct major evaluation tasks after a lengthy discussion about the activities in which they participated during their project evaluations. The discussion might have sensitized them to the extensiveness of their participation, thereby positively biasing their post-project ratings inappropriately. The results of this small study should be triangulated with the findings of additional, broader research.

This article contributes rare quantitative findings to the literature on evaluation capacity building in K–12 settings. It is another contribution to a meagre but growing body of research on building evaluation capacity. Much more remains to be done.

REFERENCES

Compton, D.W., Baizerman, M., & Stockdill, S.H. (2002). The art, craft, and science of evaluation capacity building. *New Directions in Evaluation 93*. San Francisco: Jossey–Bass.

Cousins, J.B., & Earl, L.M. (Eds.). (1995). *Participatory evaluation in education: Studies in evaluation use and organizational learning*. Washington, DC: Falmer.

Cousins, J.B., Goh, S., Clark, S., & Lee, L. (2003, June). *Integrating evaluative inquiry into the organizational culture: A review and synthesis of the knowledge base*. Paper presented at the annual meeting of the Canadian Evaluation Society, Vancouver, British Columbia.

Garet, M.S., Birman, B.F., Porter, A.C., Desimone, L., Herman, R., & Yoon, K.S. (1999). *Designing effective professional development: Lessons from the Eisenhower program*. Washington, DC: U.S. Department of Education.

Guilford, J.P. (1954). *Psychometric methods*. New York: McGraw-Hill.

Higa, T.A.F. (2004). *A study through a Vygotskian perspective of a hybrid model of participatory evaluation and school-based evaluation*. Manuscript in preparation.

Mohrmann, S.A., Wohlstetter, P., & Associates. (1994). *School-based management: Organizing for high performance.* San Francisco: Jossey–Bass.

Paul R. Brandon is Professor of Educational Evaluation at Curriculum Research & Development Group, University of Hawai'i at Manoa. He has conducted educational research, research on program evaluation, and educational/social program evaluations for nearly 30 years. He is the author or co-author of 21 refereed journal articles and approximately 150 evaluation or other technical reports. His current research and evaluation interests are evaluation capacity building, professional development, arts education, and standards of merit in program evaluation.

Terry Ann F. Higa is Evaluation Specialist at Curriculum Research & Development Group, University of Hawai'i at Manoa, and a doctoral student completing her dissertation in the University's Department of Educational Psychology. She has 12 years experience as an educational program evaluator, has authored or co-authored over 50 evaluation reports, and has served as a teacher at both the elementary and post-secondary levels. Her research interests are participatory evaluation and school-based evaluation.