
A Comparison of Participant and Practitioner Beliefs About Evaluation

American Journal of Evaluation
33(2) 208-220
© The Author(s) 2012
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/1098214011423803
http://aje.sagepub.com


Anna K. Whitehall¹, Laura G. Hill¹, and
Christian R. Koehler²

Abstract

The move to build capacity for internal evaluation is a common organizational theme in social service delivery, and in many settings, the evaluator is also the practitioner who delivers the service. The goal of the present study was to extend our limited knowledge of practitioner evaluation. Specifically, the authors examined practitioner concerns about administering pretest and posttest evaluations within the context of a multisite 7-week family strengthening program and compared those concerns with self-reported attitudes of the parents who completed evaluations. The authors found that program participants ($n = 105$) were significantly less likely to find the evaluation process intrusive, and more likely to hold positive beliefs about the evaluation process, than practitioners ($n = 140$) expected. Results of the study may address a potential barrier to effective practitioner evaluation—the belief that having to administer evaluations interferes with establishing a good relationship with program participants.

Keywords

practitioner evaluation, barriers to evaluation, evaluation anxiety

Building capacity for internal evaluation is a common organizational theme in social service delivery, education, health care delivery, community coalition work, and other disseminated services (Cullen, Giles, & Rosenthal, 2006; Fetterman & Wandersman, 2005; King, 2002; Taut, 2007). One approach to enhancing internal evaluation capacity is to designate and train internal evaluators—that is, individuals whose role within the organization or agency is to evaluate, not to deliver programs or services. A second, and increasingly popular, model is to use “practitioner evaluators”—that is, individuals whose role is both to deliver programming and to conduct evaluation (Shaw & Faulkner, 2006). Indeed, this practice is a core component of empowerment evaluation (see Fetterman & Wandersman, 2005). Practitioner evaluation may be comprehensive, including the design and administration of evaluation as well as analysis and reporting of results; it may be self-reflective and

¹Department of Human Development, Washington State University, Pullman, WA, USA

²Washington State University, Grays Harbor, WA, USA

Corresponding Author:

Laura G. Hill, Department of Human Development, Washington State University, PO Box 644852, Pullman, WA 99164, USA
Email: laurahill@wsu.edu

systematically integrated with practice; or it may be limited to executing an evaluation protocol designed by others. The skills required for a typical pretest, posttest data-gathering evaluation are likely to be different from those required for program implementation and service delivery, and interaction with clients during evaluation may have a different feel, different goals, and different pitfalls from interaction during program delivery (Mills, 2002; Myers-Walls, 2000). The goal of the present study was to extend our limited knowledge of practitioner evaluation by examining, within the context of a multisite dissemination of a family intervention, (a) practitioners' responses to conducting evaluation, (b) participants' concerns about completing an evaluation, and (c) practitioners' predictions of participants' concerns.

Benefits and Barriers

There is a substantial literature on the ways in which practitioner evaluation or research may benefit the quality of practice—for example, by encouraging self-evaluation and critical reflection about practice (Riemann, 2005; Shaw, 2005) or by helping to systematize service delivery (Cherin, 2004). Moreover, practitioner evaluation has many potential benefits over outside evaluation (Kinsey, 1981). First, practitioners can provide information about the context and operations of a program that may not be available to outside evaluators. Second, their program experience, which is deeper and more comprehensive than an outsider's, can inform both the content and the process of an evaluation and thus improve its validity and utility. Third, practitioner evaluation, if done properly, may be more practical, efficient, and cost-effective than outside evaluation, especially in ongoing or multisite evaluations. Finally, practitioner evaluation may be more collaborative and participatory than outside evaluation, by virtue of the relationship between practitioner and participants, which in turn may result in better data and ultimately improved quality of implementation (Donaldson, Gooler, & Scriven, 2002; Fetterman & Wandersman, 2005).

Understanding practitioner-directed evaluation is important because organizations and agencies that dispense funds for evidence-based programming increasingly require that programs conduct outcome evaluation, even when there is little internal capacity to do so and when resources are inadequate to hire independent evaluators. As a result, practitioners who implement programs may conduct evaluation themselves or at least collect the data. Thus, practitioner evaluation constitutes a “hidden and invisible research practice” that we have little understanding or knowledge of (Shaw & Faulkner, 2006).

At the same time, there is a common perception that practitioners do not like conducting program evaluations and some evidence to support that belief. In a survey of 224 Cooperative Extension practitioners from across the United States, Morford, Kozak, Suvedi, and Innes (2006) found that whereas 60% did not mind conducting program evaluation, 27% preferred to ignore it or dreaded it. Similarly, Kegeles, Rebhook, and Tebbetts (2005), in a qualitative study of community-based organizations (CBOs), reported that 20% of CBO leaders and two thirds of technical assistance providers felt that practitioner evaluators held negative attitudes about conducting evaluation for their programs.

Negative feelings about evaluation may be a function of practical barriers (perceived and real) to evaluation practice. For example, evaluation may feel cumbersome or intimidating to practitioners who know that evaluation should be done “right” but lack training in basic evaluation techniques (Kegeles, Rebhook, & Tebbetts, 2005; Myers-Walls, 2000; Taut & Alkin, 2003). Organizational demands for evaluation in the absence of training, technical support, and resources may engender resistance, resentment, and ultimately a negative attitude toward evaluation generally—for example, Hill and Parker (2005) found that 44% of extension faculty and staff surveyed felt that they had inadequate resources to conduct evaluation. Even with adequate training and organizational support, practitioners may perceive the time spent conducting an evaluation as time lost to program implementation (Kinsey, 1981; Morford et al., 2006; Myers-Walls, 2000), and infringements on program time are likely to foster negative responses to the process of evaluation.

Negative feelings may also be a function of motivational barriers to evaluation. For example, practitioners may be skeptical about the meaningfulness of evaluation and mistrust the reduction of rich experience to data points (Harvey & Oliver, 2002; Mills, 2002; Myers-Walls, 2000; Shaw & Faulkner, 2006), and they may be skeptical about the utility of evaluation if they do not see results used actively (Kinsey, 1981; Patton, 2008). Conversely, practitioners may fear that program evaluation *will* be used to assess their performance (King, 2002; Morford et al., 2006) or to justify discontinuing program funding (Kegeles et al., 2005). Often, even when such threats are not perceived or do not materialize, for many practitioners the process of evaluation is neither rewarding nor rewarded (Mills, 2002; Myers-Walls, 2000); thus, motivation to conduct evaluation is low.

The Present Study

In the current investigation, we examined another motivational barrier to practitioner evaluation: The possibility that evaluation may be perceived as intrusive or annoying by program participants (Hill & Betz, 2005; Myers-Walls, 2000). Even if a practitioner evaluator feels confident in her technical skills, has organizational support, and considers evaluation an integral part of program implementation, the belief that evaluation is aversive to participants could pose a significant barrier. In the present study, we explore the degree to which practitioner evaluators hold such a belief.

The Strengthening Families Program. We conducted this study within the context of an ongoing (2001–present) regional dissemination of an evidence-based family strengthening program (the Strengthening Families Program for Parents and Youth 10–14 [SFP]), delivered in many communities across the U.S. Pacific Northwest. Each SFP cycle meets once a week for 2 hr over the course of 7 weeks. During the first hour, parents and youth meet separately and participate in a combination of group activities, discussions, games, and didactic presentations. Parents learn about adolescent development, practice communicating rules clearly and warmth openly, and learn a variety of methods to provide structure while also supporting their adolescents' growing need for autonomy. Youth learn peer resistance skills, discuss their goals and dreams, and engage in activities designed to increase empathy for some of their parents' fears and concerns. During the second hour, families come together for activities designed to increase involvement and strengthen family identity. SFP has a strong research base and is considered a "model" or "exemplary" program by a number of U.S. federal agencies. Youth whose families attended the program as part of a randomized clinical trial were less likely to have initiated substance use or to abuse alcohol, tobacco, and prescription medicines than control students 6 years later; have higher school attachment and school achievement; and are less likely to report depressive symptoms (Spath, Trudeau, Gyll, Shin, & Redmond, 2009). The program is popular and is implemented widely throughout the United States as well as in Panama, Sweden, Turkey, Russia, England, Wales, Thailand, and other countries.

SFP evaluation. We provide evaluation services, free of charge, to CBOs across the Pacific Northwest region of the United States. Thus, we have extensive contact with SFP practitioners in over 40 counties. The extension system of Washington's land-grant university supports dissemination of SFP and regularly conducts trainings for practitioners from community agencies who are interested in implementing the program. Trainers present data on program effectiveness from the original clinical trial and also from the regional evaluation. They include instruction on how and when to conduct the evaluation and orient practitioners to the university's SFP website and its section on evaluation, which includes downloadable pretest and posttest forms; evaluation protocols with scripts for both parent and youth pretests and posttests; sample reports; video instructions for preparing, administering, and sending in the evaluation; and contact information for anyone who wants technical assistance with the evaluation.

The evaluation is designed to be administered by practitioners on the first night (a quantitative pretest) and last night (a quantitative posttest with several open-ended questions). Practitioners send completed evaluations to a university researcher (the second author) who, with graduate research assistants, processes the data and sends back reports about parent and youth responses to the program. The evaluation is free and in most cases voluntary—though many funders require evidence of program effectiveness, most do not require that practitioners use this particular evaluation. At many sites, evaluation reports have been useful in obtaining continued funding for the program from various agencies, foundations, or local businesses.

Over the course of 10 years, we have processed evaluation data and produced reports for over 400 occurrences of the program. The third author has delivered the program and carried out the evaluation several times, and we have had numerous discussions with other practitioners who have conducted the evaluation throughout the years. Practitioners report that the pretest and posttest each take about 15–20 min on average, though some report much longer times (e.g., 45 min). Thus, the average evaluation uses up a significant portion of the 1-hr parent or youth session on the first night and of the 1-hr family session on the seventh (last) night of the program, unless extra time is added specifically for evaluation. The process, content, and length of the evaluation have been refined over time and in response to practitioner feedback, to minimize the time demand and to make it as easy as possible to administer. Nonetheless, many practitioners are dismayed by the evaluation burden: time taken away from program content and the creation of more paperwork to process.

We continue to look for ways to minimize the practical burden of evaluation. The present study, however, arose in response to anecdotal evidence that some practitioners were reluctant to implement the evaluation not because of practical barriers but because it was perceived as intrusive by participants. Other practitioners reported that they administered the evaluation, but conducted the pretest at the second meeting of the program, after having established rapport with program participants during the first meeting. There was no evidence, however, of participants objecting to the evaluation—that is, we received no reports of participants complaining about or refusing to complete evaluations. Hence, it was unclear whether participants actually found the evaluation intrusive, or whether that was simply a perception of practitioners.

Accordingly, our main objective was to examine the match between participant attitudes and practitioner beliefs about participant attitudes. We surveyed participants who attended SFP at different sites as well as SFP practitioners throughout the state. Because we had no evidence, anecdotal or otherwise, of participants objecting to or refusing to complete the evaluation, we hypothesized that practitioner perceptions of participant attitudes would be significantly more negative than participants' self-reported attitudes. In our practitioner survey, we also included several items about specific barriers to evaluation in order to explore the applicability of findings from the studies cited above (e.g., Hill & Parker, 2005; Kegeles et al., 2005; Morford et al., 2006) in this different context.

The role of practitioner evaluators in the dissemination of SFP is participatory to some degree (e.g., in providing feedback and suggestions that lead to the ongoing revision of evaluation protocol, materials, and reporting), but the majority of practitioner involvement comes in administering the evaluation. Hence, practitioner evaluators described in the present study are not engaged in a dynamic and fully participatory evaluation process but rather in the execution of a standardized evaluation, with opportunity for input to change that evaluation.

Method

Procedure

We recruited practitioners for the study from a roster of SFP facilitators trained in the state of Washington between 2000 and 2006. After elimination of duplicate entries and those who were

missing contact information, the eligible sample comprised 326 practitioners. We conducted the survey using the tailored design method (Dillman, Smyth, & Christian, 2009), with an initial letter informing potential respondents about the survey, followed 1 week later by a letter that included the survey, consent form, a stamped envelope with the researchers' address preprinted, and a \$2.00 bill. We sent reminder postcards in Week 4 to 151 practitioners from whom we had not heard back and a final mailing in Week 6 to the remaining pool of 132 nonrespondents. Forty-three surveys were returned as undeliverable. Of the remaining 283, 124 practitioners (44% of those with valid contact information, and 38% of the total mailings, including surveys returned for invalid contact information) submitted completed surveys. An additional 16 practitioners were recruited at a practitioner training that occurred during the course of the study, resulting in a final sample size of 140.

We contacted 11 site coordinators who had previously conducted SFP to explain the study and ask if they were willing to recruit participants from upcoming programs for the participant attitude survey. All coordinators agreed to participate. At those sites, practitioners announced the 20-item survey at the beginning of the program; consenting participants completed the survey along with the standard program pretest on the first night of the program. All participants at the 11 sites completed both the pretests and the attitude survey. Surveys (and program evaluation pretests and posttests) were anonymous. Participants were given a \$5.00 gift card to a local store as a token of gratitude. The Institutional Review Board of Washington State University deemed procedures for this portion of the study exempt.

Sample Characteristics

Of the 140 practitioners who completed the survey, 76% were female. Practitioner ages ranged from 21 to 72, with an average of 49 years. All practitioners had completed at least high school or general educational development (GED). Of 105 program participants, 68% were female. Participant ages ranged from 21 to 65, with an average age of 38 years. In Table 1, we report details of practitioner and participant race and ethnicity as well as additional demographic information about practitioners. We did not collect more extensive demographic information from participants because of practitioner concern about the time required to complete surveys.

Measures

Beliefs and feelings about evaluation. We created items designed to assess positive feelings, negative feelings, and neutral feelings about evaluation. Practitioners responded to 6 items asking how they felt about evaluation (*The following questions ask you to report your own feelings about conducting program evaluations for the SFP(s) that you have facilitated*) and 10 items asking how they thought participants felt about evaluation (*Please mark the answer that best matches your thoughts about how program participants respond to being evaluated*; see Table 2 for complete list of items). The latter set of 10 items matched the 10 items that program participants answered about evaluation beliefs (*For the following questions, please mark the answer that best matches your thoughts or feelings*). Response options for both sets of questions on a Likert-type 5-point scale ranged from *strongly agree* to *strongly disagree*.

An initial exploratory factor analysis using a varimax rotation indicated that a three-factor solution best described the data for both practitioners and participants (see Table 3 for items in each factor). We then used structural equation modeling to test for adequacy of the model. Fit indices for the three-factor model indicated a good fit, $\chi^2(26) = 38.99, p = .05$; comparative fit index (CFI) = .99; root mean square error of approximation (RMSEA) = .04, 90% confidence interval (CI) = [.00, .07], whereas a single-factor solution provided an inadequate fit, $\chi^2(29) = 178.82, p < .00$; CFI = .83; RMSEA = .15, 90% CI = [.13, .17]. We averaged items within each of the three factors to create subscales; we called these subscales "positive orientation" ($\alpha = .81$) "negative orientation" ($\alpha = .79$), and "learning orientation" ($\alpha = .71$).

Table 1. Demographic Characteristics of Facilitator and Program Participant Samples From Washington State Strengthening Families Program

Demographic Characteristic	Facilitators <i>n</i> = 140 (<i>n</i> , % of total)	Program Participants <i>n</i> = 105 (<i>n</i> , % of total)
Gender		
Male	17 (12)	22 (21)
Female	107 (76)	71 (68)
Unreported	16 (11)	12 (11)
Race/ethnicity		
White/European	91 (65)	72 (67)
Asian/Pacific islander	2 (1)	0
Black/African American	3 (2)	0
American Indian/Alaska native	3 (2)	7 (7)
Latino/Latina	13 (9)	5 (5)
Other	9 (7)	5 (5)
Unreported	19 (14)	16 (15)
Age		
18–30	6 (4)	18 (17)
31–40	24 (17)	47 (45)
41–50	30 (21)	30 (29)
51–60	46 (33)	8 (8)
61–70	11 (8)	1 (<1)
71–80	1 (<1)	0
Unreported	16 (16)	1 (<1)
Questions asked of facilitators only		
Education (completed)		
Sixth grade	0	
Ninth grade	0	
Eleventh grade	0	
GED/Twelfth grade	13 (9)	
Associate's degree	21 (15)	
Bachelor's degree	44 (31)	
Master's degree	42 (30)	
PhD/professional training	4 (3)	
Unreported	16 (11)	
Employment		
Currently working	117 (84)	
Not currently working	7 (5)	
Unreported	16 (11)	
Types of jobs		
Administrator/program coordinator	40 (29)	
Counseling	32 (23)	
Parent educator	5 (4)	
Teacher/program facilitator	19 (14)	
Other	12 (9)	
Unreported	29 (21)	

Note: In some cases percentages do not sum to 100% because of rounding error.

GED = general educational development.

Marlowe–Crowne Social Desirability Scale. Program participants completed a 10-item version of the Marlowe–Crowne Social Desirability Scale (Strahan & Gerbasi, 1972); use of this scale enabled us to rule out the possibility that self-reported positive attitudes about evaluation were a function of socially desirable response bias. Sample items on the Marlowe–Crowne include “When I don’t

Table 2. Average Scale and Item Scores for Participants and Practitioners

	SFP Parent Participants (<i>n</i> = 105)	Practitioner Sample I (<i>n</i> = 16)	Practitioner Sample II (<i>n</i> = 124)
Evaluation belief scales			
Learning	3.56 (0.81)	3.34 (0.77)	3.53 (0.75)
Negative	2.52 (0.60) _a	3.17 (0.55) _b	3.04 (0.65) _b
Positive	4.23 (0.45) _a	3.56 (0.45) _b	3.51 (0.58) _b
Evaluation belief scale items—Participants and practitioner respondents ^a			
I feel that my opinions are a valuable part of the evaluation process [OR: Participants feel that their opinions are a valuable part of the evaluation process (<i>Practitioner version</i>) ^a]	4.07 (0.58) _a	3.75 (0.68)	3.56 (0.75) _b
I feel that filling out evaluations takes too much time	2.50 (0.80) _a	3.06 (1.00) _b	3.04 (0.91) _b
I answer the questions on evaluations honestly	4.56 (0.54) _a	3.06 (0.44) _b	3.40 (0.76) _b
I believe that evaluations are an invasion of my privacy	2.12 (0.79) _a	2.75 (0.77) _b	2.73 (0.79) _b
I am happy to give my opinions on evaluations	4.07 (0.77) _a	3.63 (0.72)	3.48 (0.71) _b
Some parts of the evaluation process I do not enjoy	3.07 (0.90) _a	3.63 (0.72) _b	3.55 (0.75) _b
I am uncomfortable filling out evaluations	2.35 (0.81) _a	3.25 _a (0.86) _b	2.83 (0.81) _c
I learn something about the program by completing evaluations	3.50 (0.90)	3.25 (1.00)	3.47 (0.84)
I believe that evaluating a program may help to improve it	4.33 (0.61) _a	3.81 (0.54) _b	3.62 (0.78) _b
I learn something about myself by completing evaluations	3.62 (0.87)	3.44 (0.81)	3.59 (0.3)
<i>Individual items—Practitioner respondents only</i>			
Conducting a “pretest” evaluation on the first night of a program makes it harder to establish rapport with participants	—	2.93 (1.11)	2.56 (0.96)
I am uncomfortable administering evaluation surveys to program participants	—	2.12 (0.88)	1.94 (0.85)
It is easy for me to describe to program participants why we are conducting an evaluation	—	3.81 (0.96)	3.63 (1.31)
Evaluations invade participants’ personal privacy	—	2.42 (0.82)	2.13 (0.89)
Evaluating a program may help to improve it	—	4.34 (0.81)	4.69 (0.60)
The evaluation process takes too much time	—	2.84 (1.01)	—

Note: Practitioner I = pilot sample of SFP practitioner trainees; Practitioner II = mail survey sample.

^aParticipants responded to items about how they felt (*For the following questions, please mark the answer that best matches your thoughts or feelings*); practitioners were asked to rate how they thought program participants felt (*Please mark the answer that best matches your thoughts about how program participants respond to being evaluated*), and items were slightly reworded, as shown in first items. Mean values with different letter subscripts are significantly different at $p < .05$ (Tukey’s correction used for multiple comparisons).

know something I don’t at all mind” and “I never hesitate to go out of my way to help someone in trouble” with forced-choice response options of “True” or “False.” Scale scores are calculated as the sum of “Yes” answers, so that higher scores indicate a higher likelihood of positive self-presentation. The average score in our sample was 4.54 ($SD = 2.14$, range 0–9), with a median and modal score of 4. The average scores are nearly identical to those reported by Strahan and Gerbasi

Table 3. Factor Loadings for Items on the Survey of Evaluation Beliefs

Variables	Scales		
	Negative	Positive	Learning
Uncomfortable	.80	-.31	-.07
Do not enjoy	.77	-.02	-.08
Too much time	.76	-.08	-.12
Invasion of privacy	.70	-.31	-.04
May improve program	.01	.79	.14
Answer honestly	-.19	.74	.07
Opinion valuable	-.18	.72	.24
Happy to give opinion	-.40	.61	.23
Learn about program	-.06	.20	.87
Learn about self	-.15	.20	.85
Eigenvalues	4.49	1.27	1.10404
Percent of variance explained	45%	13%	10%

Note: Sample ($n = 245$) includes both practitioners and program participants. Bolded figures represent items with high loadings on the factor represented in that column.

(1972) from a sample of college men and women (M_s 4.5 and 4.6, SD_s both 2.1), which indicates that our sample did not show an unusually strong motivation to respond in a socially desirable fashion. Reliability in our sample was low (Cronbach's $\alpha = .59$). However, this short form of the scale has been recommended by other authors because it is less burdensome and more easily accepted than the longer form, and because it is less sensitive to differences in age and socioeconomic status than other short forms (Fraboni & Cooper, 1989).

Results

Descriptive Statistics

Descriptive statistics for scale variables are shown in Table 2. For both participants and practitioners, scores on the Negative Orientation scale are negatively correlated with those on the Positive and Learning Orientation scales (see Table 4). Scores on the short version of the Marlowe–Crowne Social Desirability scale, administered only to program participants, were not significantly correlated with any of the Orientation scales, indicating that responses to those scales were not biased by a motivation to provide socially desirable answers.

Practitioner Perceptions of Participant Orientation Toward Evaluation

Results confirmed our hypothesis that practitioners would overestimate participants' negative feelings and underestimate their positive feelings about evaluation. Independent t tests showed that practitioners rated participants lower in their Positive Orientation ($t = 11.38, p < .001$) and significantly higher in their Negative Orientation ($t = 6.70, p < .001$) toward evaluation than participants rated themselves. We had no specific hypothesis about Learning Orientation, and practitioners and participants did not differ in their response to the Learning Orientation items ($t = 0.54, p = .59$).

Practitioners' Own Beliefs and Feelings About Evaluation

Almost 30% of practitioners felt that administering an evaluation on the first night of a program made it harder to establish rapport with participants, but only 9% felt that evaluation was an invasion of participants' privacy. Only 8% felt uncomfortable administering an evaluation; 76% said it was easy for them to explain the reasons for conducting an evaluation; and 93% felt that evaluation might

Table 4. Correlations of Evaluation Beliefs Subscales for Practitioners and Participants

	Negative Orientation	Positive Orientation	Learning Orientation	Missingness
Negative orientation	—	-.46*	-.27*	.06
Positive orientation	-.61*	—	.45*	.00
Learning orientation	-.53*	.49*	—	-.10
Missingness	.04	.01	-.02	—

Note: Intercorrelations of subscales above the diagonal represent the practitioner sample ($n = 140$) and below the diagonal represent the program participant sample ($n = 105$).

* $p < .05$.

help improve a program. Twenty-five percent felt that evaluation took up too much time. In short, practitioner attitudes about evaluation were largely positive, though a minority believed that it interfered with their initial relationship building with participants and were concerned about the time burden.

Post Hoc Analyses

We conducted post hoc regression analyses to determine whether practitioner characteristics, including their own feelings about evaluation, were related to their estimates of participant feelings about evaluation. We controlled for practitioners' own beliefs about evaluation and examined the relation of practitioner demographics (gender, minority status, education, and age) with their beliefs about participant attitudes. Older practitioners were significantly more likely to believe that participants felt positively about evaluation ($B = 0.02$; $p < .001$); demographic characteristics were unrelated to negative or learning beliefs. Among participants, demographic characteristics were unrelated to scores on any of the belief subscales.

We also examined whether the number of missing values (i.e., items with no response) was associated with beliefs about evaluation. More than 80% of respondents answered all items. The proportion of program participants with complete data (84.8%) was higher than that of practitioners (74.6%), and a t test of mean differences in amount of missing data was significant ($t = 3.13$, $p < .005$). The correlation of amount of missing data with positive, negative, and learning belief subscales was nonsignificant among both participants and practitioners, indicating that specific beliefs about evaluation, as measured by our items, did not influence willingness to complete all items on the evaluation.

We also compared the participants in the current study with all other SFP participants who have participated in the regional SFP evaluation between 2003 and 2011 ($n = 4,911$) to see if there were systematic differences between the two groups. We found no significant differences between participants in the current study and SFP participants in age ($t = .94$, $p = .35$) or gender ($\chi^2 = 1.77$, $p = .18$).

Discussion

The goal of this study was to examine one possible motivational barrier to conducting program evaluation. Specifically, we hypothesized that practitioners would estimate that participants in a family-strengthening intervention would hold more strongly negative and less strongly positive attitudes about evaluation than they actually did. Results supported this hypothesis: participants were significantly more positive and less negative than practitioners thought they were. Examination of practitioners' own feelings about evaluation revealed that only a minority believed that evaluation was intrusive, and a strong majority felt confident about administering the evaluation and believed it was useful. Consistent with previous research, about a quarter to a third of practitioner evaluators felt that

evaluation took too much time and interfered with initial rapport building. However, to the best of our knowledge, this is the first study to demonstrate that practitioner attitudes may differ from those of the groups they serve (but see Schwandt & Dahler-Larsen, 2006).

Limitations

Our findings and their implications for training are likely influenced by the context of the evaluation. The program in question, SFP, is a voluntary intervention with an emphasis on promoting family strengths; this positive focus may explain in part the positive attitudes of participants toward evaluation. It is a universal program and therefore does not target families identified as having problems, nor does the evaluation ask parents to report highly sensitive information (e.g., about physical discipline or substance use). Findings may not generalize to programs with different types of content (e.g., remediation or court-mandated programs), different types of clientele (e.g., participants with legal difficulties or those with physical or emotional vulnerabilities), or different types of evaluation (e.g., those that request sensitive information).

All practitioner evaluators who were asked to have their programs take part in the study of participant attitudes agreed, and all participants in those programs present at pretest completed the survey. However, the practitioner survey mailed to all eligible practitioners from the SFP training roster had a low response rate (44%), which raises the possibility of response bias. Of particular concern is the bias we would encounter if practitioners in our sample were unusually negative about evaluation. In that case, we would have a sample of program participants who were (conceivably) more positive than most about evaluation because of the nature of the program compared to a sample of practitioners who were more negative about evaluation. This would exaggerate the difference between practitioner and participant beliefs, thus limiting the generalizability of our findings and the likelihood of their replication even in similar settings.

Three factors argue against a strong concern about a negativity bias: First, our findings on practitioner reports of perceived barriers to evaluation are congruent with those of other studies (e.g., similar proportions of practitioner evaluators concerned about evaluation taking up too much time, as reported in Kinsey, 1981; Morford et al., 2006; Myers-Walls, 2000). This leads us to believe that concerns about program evaluation interfering with programming or with the establishment of rapport may be common to practitioners across a variety of settings. Second, the subsample of practitioners whose programs participated in the study were not significantly different in their beliefs, or their estimates of participant beliefs, than the overall sample of practitioners who completed the mailed survey. We believe that the former group, who voluntarily committed to their programs' involvement in the study, were likely to represent the positive end of the spectrum of beliefs about evaluation. Since they were no different from the larger group, it does not appear that practitioner responses were especially negatively biased. And finally, we believe that practitioner evaluators who held especially negative attitudes about evaluation would be less likely to complete and return the survey, and that our sample therefore represents those who are more positive about evaluation generally and presumably also in their estimates of participant attitudes. This assumption is also consistent with the attitudinal match between the smaller and larger practitioner subsamples. Thus, it appears that our findings are in fact conservative: With a more representative sample, there would have been an even greater discrepancy between practitioner estimates and participant self-reports.

However, our inferences are speculative and must be weighed in light of the possibility of selection bias. In addition, the generalizability (and meaning) of these findings can only be established with additional research. Future studies should include representative practitioner and participant samples from different types of programs and with a sample that allows for comparison of respondents with nonrespondents.

Implications for Practice

These findings have implications both for program practice and for the training of practitioners and others, such as teachers, who may be required to conduct evaluations of programs and curricula delivered in community or school settings. An unexpected finding was that both practitioners and program participants saw evaluation as a learning opportunity. This finding is consistent with principles of both empowerment evaluation (Fetterman & Wandersman, 2005) and organizational development (Preskill & Torres, 1999)—ownership of the evaluation process enhances credibility of findings and creates a fertile context for learning. When administration of an evaluation is a requirement of program delivery, it may be helpful to emphasize in training for programs such as the one described here the fact that participants generally hold positive attitudes about evaluation, do not find it intrusive, and understand its value as a quality improvement tool. (We reemphasize, however, that this finding may not generalize to other types of programs.) In other words, the data presented here may help dispel practitioner concerns about negative effects of administering pretests and posttests.

In addition, training of practitioner evaluators can include discussion of the utility of evaluation as part of programming itself, aside from quality improvement or documentation of outcomes (Taut, 2007). For example, pretests may provide an opportunity to preview program content and goals, and posttests provide an opportunity to reflect on what has been learned (Hill & Betz, 2005). We now use this information when we present the evaluation in practitioner trainings, proposing that the evaluation be approached (and also discussed with participants) as contributing to program goals rather than simply as an external imposition. Indeed, practitioner evaluators are in a much better position to integrate evaluation into program practice than outside evaluators. This provides an additional rationale for the participation of practitioner evaluators in the design of an evaluation since alignment of evaluation content and process with program practice would increase the likelihood of its utility, usability, and subsequent integration into program practice. By extension, and in the tradition of action research, input from participants on how evaluation contributes to their experience of a program would increase the validity and utility of evaluation as a component of practice.

Perhaps more meaningfully, the finding that participants and practitioners both perceived evaluation as an opportunity to learn more about the program and about themselves suggests richer possibilities for practitioner evaluation and research. In the present study, the evaluation was a traditional, standardized pretest/posttest tool used primarily to document quantitative program outcomes for funders, a format often experienced by practitioners as intrusive and external to their service goals. However, knowing that participants appreciate evaluation as a mode of learning might help practitioners think about how to engage participants in a more reflective evaluation process—to pose questions that would facilitate participant learning as well as their own and that would provide information directly useful in improving practice.

Further Research and Conclusions

The present study was designed to answer a single question that arose directly from practice, in the context of an ongoing program evaluation: Did program participants feel as negatively about evaluation as many practitioner evaluators appeared to think? We approached the study, and have presented it, as a straightforward empirical examination of that particular question. However, design of future research on practitioner evaluation in evidence-based program delivery would benefit from attention to the larger theoretical literature on practitioner research. In particular, future studies examining barriers to evaluation and resistance to conducting evaluation, reviewed briefly in the introduction, might be expanded to include questions about whether the content, method, and results of evaluation are perceived as congruent with professional practice and identity; and about

how those perceptions vary as a function of the extent to which practitioners experience ownership of the evaluation (Fook, 2002; Lang, 1994; Lunt et al., 2008).

One significant concern, reported by nearly a third of practitioner evaluators, was that evaluation on the first night of a program interferes with their ability to build rapport with participants. Data in the current study cannot tell us the extent to which this concern is warranted from the perspective of participants; an observational study, direct survey of participant engagement, or interviews of participants would be needed to provide information about the relation of pretest evaluation to rapport building. In addition, participant attitudes about evaluation should be explored in other, more sensitive, contexts (e.g., programs for substance abuse treatment or domestic violence), where questions are more likely to be experienced as intrusive.

In conclusion, our results provide information about a specific barrier to practitioner evaluation—the concern that participants do not like to complete program evaluations—within the context of a universal, community-based family intervention. Our hope is that by providing this information to practitioner evaluators and trainees, we will decrease one source of evaluation anxiety (Donaldson et al., 2002) and stimulate additional research into practitioner evaluation.

Acknowledgment

The authors would like to thank Brianne Hood for her research assistance, the program facilitators and families who participated in the program evaluation, and the Washington state Department of Social and Health Services for their collaboration.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: The study was supported in part by two grants from the National Institute of Drug Abuse of the U.S. National Institutes of Health (R21 DA025139-01A1 and R21 DA19758-01).

References

- Cherin, D. (2004). Collecting and using data for organizational learning. In M. J. Austin & K. M. Hopkins (Eds.), *Supervision as collaboration in the human services: Building a learning culture* (pp. 240-251). Thousand Oaks, CA: Sage.
- Cullen, A., Giles, T., & Rosenthal, J. (2006). *Evaluating community-based child health promotion programs: A snapshot of strategies and methods*. Portland, ME: National Academy for State Health Policy.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). Hoboken, NJ: John Wiley.
- Donaldson, S. I., Gooler, L. E., & Scriven, M. (2002). Strategies for managing evaluation anxiety: Toward a psychology of program evaluation. *American Journal of Evaluation, 23*, 261-273. doi:10.1016/s1098-2140(02)00209-6
- Fetterman, D. M., & Wandersman, A. (2005). *Empowerment evaluation principles in practice*. New York, NY: Guilford Press.
- Fook, J. (2002). Theorizing from practice: Towards an inclusive approach for social work research. *Qualitative Social Work, 1*, 79-95.
- Fraboni, M., & Cooper, D. (1989). Further validation of three short forms of the Marlowe-Crowne scale of social desirability. *Psychological Reports, 65*, 595-600.
- Harvey, J., & Oliver, M. (2002). Towards effective practitioner evaluation: An exploration of issues relating to skills, motivation and evidence. *Educational Technology & Society, 5*, 1-11.

- Hill, L. G., & Betz, D. L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation, 26*, 501-517.
- Hill, L. G., & Parker, L. A. (2005). Extension as a delivery system for prevention programming: Capacity, barriers, and opportunities. *Journal of Extension, 43*. Retrieved from <http://www.joe.org/joe/2005february/a1.shtml>
- Kegeles, S. M., Rebchook, G. M., & Tebbetts, S. (2005). Challenges and facilitators to building program evaluation capacity among community-based organizations. *AIDS Education and Prevention. Special Issue: Building HIV Prevention Capacity through Community Collaborative Research, 17*, 284-299. doi:10.1521/aeap.2005.17.4.284
- King, J. A. (2002). Building the evaluation capacity of a school district. *New Directions for Evaluation, 93*, 63-80.
- Kinsey, D. C. (1981). Participatory evaluation in adult and nonformal education. *Adult Education Quarterly, 31*, 155-168.
- Lang, N. (1994). Integrating the data processing of qualitative research and social work practice to advance the practitioner as knowledge builder: Tools for knowing and doing. In E. Sherman & W. Reid (Eds.), *Qualitative research in social work* (pp. 265-270). New York, NY: Columbia University Press.
- Lunt, N., Fouché, C., & Yates, D. (2008). "Growing research in practice": Report of an innovative partnership model. Wellington, Australia: Families Commission.
- Mills, D. S. (2002). Examining the structural and philosophical positions of community-based organization practitioners. *Dissertation Abstracts International, 62*(9-B), 4620.
- Morford, S., Kozak, R., Suvedi, M., & Innes, J. (2006). Factors affecting program evaluation behaviours of natural resource extension practitioners: Motivation and capacity building. *Journal of Extension, 44*. Retrieved from <http://www.joe.org/joe/2006june/a7.php>
- Myers-Walls, J. A. (2000). An odd couple with promise: Researchers and practitioners in evaluation settings. *Family Relations, 49*, 341-347. doi:10.1111/j.1741-3729.2000.00341.x
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.
- Preskill, H., & Torres, R. T. (1999). Building capacity for organizational learning through evaluative inquiry. *Evaluation, 5*, 42-60.
- Riemann, G. (2005). Ethnographies of practice—Practicing ethnography. *Journal of Social Work Practice, 19*, 87-101.
- Schwandt, T. A., & Dahler-Larsen, P. (2006). When evaluation meets the "rough ground" in communities. *Evaluation, 12*, 474-495.
- Shaw, I. (2005). Practitioner research: Evidence or critique? *British Journal of Social Work, 35*, 1231-1248.
- Shaw, I., & Faulkner, A. (2006). Practitioner evaluation at work. *American Journal of Evaluation, 27*, 44-63. doi:10.1177/1098214005284968
- Spoth, R., Trudeau, L., Gyll, M., Shin, C., & Redmond, C. (2009). Universal intervention effects on substance use among young adults mediated by delayed adolescent substance initiation. *Journal of Consulting and Clinical Psychology, 77*, 620-632.
- Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlowe-Crowne social desirability scale. *Journal of Clinical Psychology, 28*, 191-193. doi:10.1002/1097-4679(197204)28:2<191::AID-JCLP2270280220>3.0.CO;2-G
- Taut, S. (2007). Studying self-evaluation capacity building in a large international development organization. *American Journal of Evaluation, 28*, 45.
- Taut, S. M., & Alkin, M. C. (2003). Program staff perceptions of barriers to evaluation implementation. *American Journal of Evaluation, 24*, 213-226. doi:10.1016/s1098-2140(03)00028-6