
Assessing Implementation Integrity of a National Nutrition Education Program: A Case Study of Share Our Strength's Operation Frontline

Chris L. S. Coryn
Western Michigan University

Daniela C. Schröter
The Evaluation Center, Western Michigan University

Anne Cullen
Fetzer Institute

Laura Seman
Share Our Strength, Cooking Matters

Janet McLaughlin
Share Our Strength, Cooking Matters

Background: Treatment implementation is not just one thing but rather is a multifaceted process that includes treatment delivery, treatment receipt, and treatment adherence. As such, local variations in implementation and service delivery of interventions are an inevitable.

Purpose: To assess implementation fidelity of a multi-site experiential nutrition education program.

Setting: Multiple sites throughout the continental United States.

Intervention: An experiential nutrition education program.

Research Design: A concurrent mixed methods design was used to assess implementation fidelity.

Data Collection and Analysis: Multiple methods of data collection and analysis were used including observations, interviews, survey questionnaires, and extant data.

Findings: Although implementation fidelity varied over program sites, overall implementation fidelity was very good and when it varied, it varied to local site needs and context.

Keywords: *program fidelity; program implementation; nutrition education*

Journal of MultiDisciplinary Evaluation
Volume 8, Issue 19

JMDE
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180
<http://www.jmde.com>

Evaluation studies across numerous substantive areas, particularly in education and health, have increasingly investigated the quality of program delivery in an effort to learn about and understand the successes and failures of applied interventions (Davis, Baranowski, Resnicow, et al., 2000; Reynolds, Franklin, Leviton, et al., 2000). In the simplest case, process evaluation is the assessment of everything that occurs prior to the emergence of true outcomes, in particular design (i.e., the degree to which program inputs and activities are logically, plausibly, empirically, or otherwise coupled to anticipated outcomes and effects) and implementation (i.e., the degree to which an intervention is executed as planned) (Coryn, 2007; Davidson, 2005; Scriven, 2007). When coupled with high-quality outcome evaluation, process evaluation offers numerous practical advantages over the typical goal-achievement and black box models of evaluation, which consider only the ends and not the means by which ends are supposedly achieved (Coryn, Noakes, Westine, & Schröter, 2011). Such evaluations not only allow stakeholder groups to gauge and diagnose areas of improvement, such as large variances in program implementation, which might result in dose-response patterns in observed outcomes, for example, but also failures to logically or empirically connect intervention strategies to the actual needs of program consumers and impactees (Coryn, Gugiu, Davidson, & Schröter, 2008). Such assessments can, for instance, provide information intended to explain how or why anticipated results occur or fail to occur, if such are plausibly attributable to design or implementation failures (Coryn, Schröter, & Hanssen, 2009).

While process evaluations of complex, multi-site interventions can pose serious practical and methodological challenges, they can also provide numerous advantages to program implementation, administration, and long-term sustainability. Operation Frontline (OFL; recently renamed Cooking Matters) is an experientially-based, participatory nutrition education program which seeks to increase food security, healthy eating behaviors, and improve the food environment for low-income families with children.

Case Example: Operation Frontline

OFL seeks to achieve these ends through precisely prescribed instructional and curricular activities and strategies implemented by professional chef and certified nutrition volunteers (the OFL program is described in detail by Swindle, Baker, & Auld, 2007). Before entering into a 3-year

period of rapid expansion in which OFL doubled its partner sites from 15 to 30 communities across the United States, OFL sought to better understand the quality of program implementation across its partners, as well the value of its program processes through conducting a process evaluation. As such, the sheer size of the program itself poses serious practical constraints (e.g., financial resources, human resources, time) related to conducting a comprehensive process evaluation that is simultaneously useful, actionable, cost-effective, timely, and valid.

Aims and Objectives

The aims and objectives of this investigation included, but were not limited to:

1. To what extent does OFL program implementation at partner sites comport with the design of the OFL program?
2. To what extent does variability in OFL program implementation at partner sites explain variability in program outcomes?
3. What factors inhibit and enable high quality OFL program implementation at partner sites?

Method

Research Design

The research design used to assess implementation fidelity of OFL across a sample of partner sites, and to extrapolate (i.e., generalize) to other sites, was a concurrent mixed methods design (i.e., qualitative and qualitative methods occurred simultaneously). As a mixed methods study, a multitude of both quantitative and qualitative data collection strategies were utilized, including observational checklists, semi-structured interviews, survey questionnaires, and extant data. The rationale for using a mixed methods approach for the study was that any mono-method biases associated with individual methods would be greatly reduced. From an evaluation-theoretic perspective, the study is best classified as collaborative and participatory, given that OFL staff were significantly involved in scoping and framing the evaluation, formulating and prioritizing evaluation questions, developing instruments and measurement devices, and interpreting results (Brandon, 1998; Cousins & Whitmore, 1998; Cullen, 2009; Cullen, Coryn, & Rugh, 2011). Even though large amounts of

quantitative data were gathered, the study is largely qualitative.

Sample

A nonprobability, purposive sample of 6 OFL partner sites were selected for observations and interviews with program staff and participants. The sites were selected using maximum variation sampling (Patton, 2002) from which inclusion was based on purposefully selecting sites with a wide range of variation on numerous dimensions of interest such as geographic location, type of implementing agency, number of OFL courses offered annually, proportion of participants completing courses, and length of time implementing OFL, for example. Due to time and resource constraints, obtaining a larger, statistically representative sample of sites was unfeasible. At each partner site, OFL courses were observed and rated by two independent observers. Interviews were conducted with OFL staff, chef and nutrition instructors, and other volunteers, as well as with a selection of course participants at each site. Additionally, survey questionnaires were administered to program staff at all OFL partner sites and, lastly, extant data for all sites were obtained from OFL headquarters in Washington, DC.

Instrumentation and Measures

Observation Checklist. The observation checklist consisted of 23 items designed to measure instructional and curriculum fidelity as set forth in *OFL Implementation Guidelines* (2009) and curricular materials. Each of the 23 items (e.g., “Participants are encouraged to discuss solutions to shared challenges around making behavior changes,” “Participants are instructed in and practice safely handling and preparing foods”) were rated by observers either as 0, 1, 2, or not applicable, where 0 = never, 1 = sometimes, and 2 = always. This simple three-point rating scale was intentionally selected in order to increase interrater reliability (but at the cost of a small loss of detail) and also because the observational checklist was designed to be used more widely by OFL staff and administrators as a monitoring tool following completion of the current study. Therefore, it needed to be simple and straightforward in its application. To further improve interrater reliability and rater calibration, exemplars of poor- and high-quality implementation accompanied each of the 23 items. Observers attended and independently rated a

total of 16 courses at 6 sites. Interrater agreement, estimated as a coefficient of agreement represented by the total proportion of observations (p_o) of which there was agreement, or

$$P_o = \frac{\text{number of exact agreements}}{\text{number of possible agreements}} = \frac{\sum f_o}{N},$$

where $\sum f_o$ is the sum of the frequencies of observed agreements, and N is the number of pairs of scores, for the rater pairs was $p_o = .80$. In addition, a single-facet generalizability (G) study was conducted to more fully examine sources of variance in the measurement procedure with raters as a random facet. Variability in obtained ratings due to rater differences would be undesirable and contribute to unreliability. The formula used to estimate the generalizability coefficient was

$$\hat{\rho}_i^2 = \frac{\sigma_p^2}{\sigma_p^2 + (\sigma_i^2 + \sigma_e^2)/n_i}$$

The variance estimates for the G-study were derived from the results of an analysis of variance (ANOVA). The ANOVA yielded three distinct sources of variance necessary for estimating the generalizability coefficient: raters, targets, and raters \times targets. The interaction term, raters \times targets, contains several sources of variance that cannot be separated, including information about random, unaccounted for variance (i.e., the residual error term). The rater component indicates systematic and overall differences in the way that raters rated classes. If one rater simply rated consistently higher than another then that would show in the rater component of variance. The target component reflects real differences in the classes.

The sources of variance described above yielded a generalizability coefficient of $\hat{\rho}_i^2 = .71$. Of the percentages of variance in scores that are attributable to the different sources of variance, 74% of the variability in scores could be attributed to the targets (this variability is desirable), whereas less than 20% could be attributed to raters (e.g., systematic differences in ratings; undesirable) and 6% to error, or the interaction between raters and targets.

Internal consistency for the 23-item measure of observed fidelity was Cronbach's $\alpha = .87$. An index of observed fidelity was computed by averaging rater-pair's scores over all observations for each site. Throughout the remainder of this

paper this construct is referred to as “index of observed fidelity.”

Semi-Structured Interviews. Semi-structured interviews were conducted with OFL staff, chef instructors, nutrition instructors, and other volunteers, as well as with a sample of course participants at each of the sites. Interviews with OFL staff sought to investigate methods of and barriers to implementation. Interviews with OFL volunteers were designed to gather information regarding their perceptions of the quality of training received to deliver course content and their teaching competencies, among others. Interviews with samples of participants were intended to elicit information about their experience and satisfaction with the course, knowledge gained by participating in the course, and application of that knowledge outside of the classroom, for example. Semi-structured interviews were conducted to gather richer data than is possible with more formalized, fixed measurement devices and because of their flexibility for investigating phenomena as they arise during the interview process.

In all, 9 OFL staff, 3 AmeriCorps volunteers, 6 professional chefs, 3 home chefs, 4 professional nutritionists, 4 nutrition students, 3 other volunteers, and 31 adult participants were interviewed during site visits.

Questionnaires. Web-based survey questionnaires were administered to all OFL partner sites (total $N = 37$). Respondents included OFL administrators ($n = 17$), OFL staff ($n = 9$), and AmeriCorps members serving at OFL partner sites ($n = 11$). Questionnaires were divided into seven major sections and consisted of both open- and closed-ended items designed to tap into a variety of implementation- and process-related issues. The first section asked questions such as number and type of full- and part-time program staff, financial resources to operate the local OFL program, and sources and amounts of funding to support local implementation, among many others. The second section asked questions about local site partners. The third section asked questions about volunteer recruitment, training, and retention. The fourth section asked questions about courses and course implementation. The fifth section asked questions about participant recruitment and retention. The sixth section asked questions about use and quality of OFL resources such as training materials and *Implementation Guidelines*, for example. The final section asked respondents to self-assess the quality of implementation at their local site. In all, administrators were asked a total of 55 questions,

not including several multi-item scales, and staff received 51 questions (excluding questions about type of implementing agency, year of first implementation, estimated annual OFL program budget in the last fiscal year, and funding sources). AmeriCorps members were asked 19 questions targeted at their specific concerns, yet included questions about OFL courses and their implementation that were also asked to the administrators and staff. Of item clusters that conceptually lent themselves to reliability analyses, internal consistency for a 14-item measure of participant experience as designed was Cronbach's $\alpha = .73$, $.78$, and $.75$ for administrators, staff, and AmeriCorps, respectively. For a 13-item measure of instructional delivery, Cronbach's $\alpha = .82$, $.95$, and $.88$ for administrators, staff, and AmeriCorps, respectively. Both of these multi-item scales were measured using 5-point Likert-type response sets. Indexes of perceived participant experience as designed and perceived instructional delivery as designed were computed by averaging across their respective item sets. Throughout the remainder of this paper, these constructs are referred to as “index of perceived participant experience” and “index of perceived instructional delivery,” respectively. Where more than one response per OFL site was obtained (e.g., multiple direct service staff), an average response was used in the analysis.

Extant Data. OFL headquarters in Washington, DC frequently gathers information from their host/partner sites for a variety of purposes (e.g., monitoring, accountability). Largely, these data are collected using Web-based survey methods and include number and type of OFL courses offered in the past year, number of participants attending each course, and proportion of participants completing each course, among many others. For this particular study, only fiscal year 2008 data were used.

Results

Observed Fidelity

Means and standard deviations for each of the 23 items and the composite index of observed fidelity are shown in Table 1. The index of observed fidelity, representing the average rating across all 23 items, was $M = 1.20$ ($SD = 0.71$).

Table 1
Means and Standard Deviations for Observed Fidelity

Item	M	SD
1. Team-teaching opportunities between chefs and nutritionists are utilized.	0.73	0.70
2. Opportunities to engage participants in discussion and activities are utilized.	1.33	0.48
3. Participants engage in food preparation.	1.63	0.49
4. Participants are instructed in and practice safely handling and preparing foods.	1.17	0.64
5. Participants are instructed in and practice food resource management in class.	1.14	1.03
6. Food made in class is consumed in a communal manner.	1.50	0.66
7. Nutrition, cooking, and budgeting activities are synchronized to make use of time available.	0.91	0.81
8. Participants are encouraged to discuss solutions to shared challenges around making behavior changes.	0.65	0.93
9. Instructor keeps class on schedule.	1.17	0.70
10. Instructor manages discussions that are off-topic.	0.90	0.85
11. Instructor reinforces key concepts.	1.45	0.74
12. Instructor reinforces key skills.	1.05	0.90
13. Instructor discusses informational handouts and course materials.	1.38	0.65
14. Instructor integrates participant's experiences into lessons.	1.17	0.76
15. Instructor encourages participants to prepare healthy foods at home.	1.54	0.51
16. Instructor encourages participants to prepare low-cost foods at home.	1.05	0.95
17. Instructor manages inappropriate behaviors.	0.81	0.91
18. Instructor builds a welcoming atmosphere.	1.75	0.53
19. Instructor demonstrates knowledge of subject-matter being taught.	1.58	0.65
20. Instructor facilitates and focuses group conversation around lesson topics.	1.04	0.86
21. Instructor builds upon previous lesson content.	1.55	0.69
22. Instructor arrives prepared and organized to teach lessons.	1.58	0.65
23. Instructor demonstrates respect for participants and their life experiences.	1.75	0.53
Index of Observed Fidelity	1.20	0.71

Some administrators pointed out that building on concepts and activities that other instructors present, assessing whether participants understood previous lesson content, and keeping participants on task by managing discussions that are "off-topic" are rarely observable

Moreover, the degree to which co-instructors (e.g., chefs and nutritionists) interacted and collaborated had a direct impact on the quality of

implementation. In many of the classes observed, there was a high-level collaboration and interaction between instructors. Specifically, many instructors used team teaching skills by (1) asking each to contribute to the discussion based on his/her knowledge/expertise, (2) reinforcing key points each other had made, and (3) weaving the cooking and nutrition education components together. However, in other instances, instructors worked in isolation and rarely interacted when presenting OFL material. This was not necessarily a deficit on the part of the instructors; rather, volunteers had been recruited at the last minute and did not have sufficient time in which to prepare for that day's session. Moreover, if instructors did not arrive early enough or did not communicate via phone or email before the class to allow discussion and preparation with one another, there was a clear reduction in the quality of implementation. While some volunteers had observably better instructional skills, either through training by OFL staff members or experience, others did not. Therefore, it is critical that these instructors be given extra time and coaching on how to collaborate with their co-instructors.

In terms of course implementation, the level of cultural sensitivity and respect was one of the biggest strengths observed in every class observed. OFL classes serve a wide range of individuals with diverse ethnicities, ages, and backgrounds. Each participant is treated with respect for his/her individual circumstance.

Relationships Between Program Inputs, Program Outputs, Index of Perceived Participant Experience, and Index of Perceived Instructional Delivery

Means and standard deviations for program inputs and outputs (including number of classes offered, number of participants, number of participants completing classes, and graduation rate) from extant OFL data, and administrator and staff indices of perceived participant experience and perceived instructional delivery are shown in Table 2.

Table 2
Means and Standard Deviations for Program Outputs,
Index of Perceived Participant Experience, and Index of
Perceived Instructional Delivery

Variable	<i>M</i>	<i>SD</i>
Number of classes	28.67	16.59
Number of participants	345.22	205.68
Number of graduates	275.22	173.22
Graduation rate	79.87%	10.62%
Administration perceived participant experience	4.19	0.31
Administration perceived instructional delivery	3.87	0.29
Staff perceived participant experience	4.15	0.12
Staff perceived instructional delivery	4.25	0.25

If administrators and staff report that participants experience the program as designed and that instruction is delivered according to prescribed guidelines, then it would also be anticipated that as the quality of participant experience and instructional delivery increases so to should program outputs (in particular the number of graduates and graduation rates). That is, if implementation is of high quality then, logically, participants would be more likely to remain in the program and less likely to drop out. For administrator perceptions of participant experience and instructional delivery, however,

only modest correlations, $\rho = 0.20$ and $\rho = -0.37$, between number of graduates and perceived participant experience and instructional delivery, respectively, were found (see Table 3). For participant graduation rate and administrator perceptions of participant experience and instructional delivery, these correlations too were small in magnitude ($\rho = -0.07$ and $\rho = -0.24$ for participant experience and instructional delivery). For staff, these correlations were $\rho = 0.28$ and $\rho = 0.29$, and $\rho = 0.01$ and $\rho = 0.10$, for number of graduates and participant graduation rate correlated with their perceptions of participant experience and instructional delivery, in that order. Correlations between administrator and staff perceptions of participant experience and instructional delivery were $\rho = -0.20$ and $\rho = 0.06$, respectively, which are small in magnitude and in one instance negative in direction. If these correlations were large in magnitude and positive in direction, then there would be some evidence for cross-validation. These correlations, being small in magnitude and negative in direction, however, can plausibly be interpreted as inconsistencies in perceptions between administration and staff as to the quality of participant experience and instructional delivery.

Table 3
Correlations Between Program Outputs, Index of Perceived Participant Experience, and Index of Perceived Instructional Delivery

Variable	1	2	3	4	5	6	7
1. Number of classes							
2. Number of participants	0.97						
3. Number of graduates	0.93	0.97					
4. Graduation rate	-0.02	-0.03	0.20				
5. Administration perceived participant experience	0.24	0.27	0.20	-0.37			
6. Administration perceived instructional delivery	-0.11	-0.04	-0.07	-0.24	0.34		
7. Staff perceived participant experience	0.06	0.19	0.28	0.29	-0.20	0.03	
8. Staff perceived instructional delivery	-0.14	-0.05	0.01	0.10	0.06	0.06	0.59

Note: Significance tests (i.e., *p*-values) are not reported given that the correlations are population parameters (i.e., ρ) rather than sample statistics (i.e., *r*).

Antecedents of Process-Related Outputs

An ordinary least squares (OLS) multiple regression analysis was used to investigate the extent to which five program inputs (number of years operating an OFL program, monetary resources, in-kind resources, number of chef volunteers, and number of nutritionist volunteers)

predicted the immediate process-related output of graduation rate. Combined, the five predictor variables explained 68% (adjusted $R^2 = .68$) of the variability in the criterion; $F(5, 16) = 11.368$, $p < .001$. As shown in Table 4, three of the five predictors were statistically significant, with the greatest weight given to number of chef volunteers ($\beta = .57$), followed by number of nutritionist

volunteers ($\beta = .55$), and then by number of years operating an OFL program ($\beta = .35$). Neither financial resources nor in-kind resources were

statistically significant predictors of graduation rate.

Table 4
Antecedents of Participant Graduation Rates

Predictor	<i>b</i>	<i>SE b</i>	β	<i>p</i>
Years operating an OFL program	1.172	.511	.345	.033
Monetary resources	.000	.000	.313	.062
In-kind resources	-.000	.000	-.006	.965
Number of chef volunteers	.422	.104	.569	.000
Number of nutritionist volunteers	.584	.141	.546	.001

Both outputs and antecedent inputs are process variables that occur prior to true outcomes, and empirically understanding the degree to which inputs affect outputs can provide valuable improvement-related information, such as where to increase particular activities or strategic efforts (recruiting chef and nutritionist volunteers) and to identify and obtain resources to sustain existing program providers (Schröter, 2008), for example. Without graduates, proximal and distal outcomes such as improved knowledge, skills, attitudes, and other abilities related nutrition behaviors and health outcomes simply cannot occur.

Additional Insights from Observations, Interviews, and Questionnaires

Effectively and efficiently implementing the OFL curricula requires a great deal of resources and creativity, which often must be stretched beyond available resource limits. While OFL provides a wide range of resources to its local program partners, including curriculum and program materials, evaluation services, training and technical assistance, AmeriCorps members, funding, networking opportunities, and other support, findings indicate that locally available resources vary tremendously for different locations. Budgets for programming range from as low as \$30,000 to a high of \$370,000 (including both monetary and in-kind support). This leads to some programs having several program staff fully committed to OFL programming exclusively, whereas others juggle varying programs and commitments in addition to OFL.

During site visits, last minute cancellations from planned volunteers resulted in replacement volunteers being recruited the night before a class or OFL staff members canceling personal plans to fill in. While these solutions work in the short term, they are not sustainable on a long-term

basis. Most importantly, not having professional chefs is a clear deviation from the protocol established in the *OFL Implementation Guidelines*. While last minute cancellations due to emergencies and unforeseen events are certainly understandable, when they become the norm, rather than the exception, they represent a clear challenge.

Moreover, the implementation of OFL requires a great deal of coordination and organization to arrange the scheduling of classes, volunteer recruitment, and purchasing of groceries, among others. To mitigate these constraints, program partners need to be well organized to ensure successful coordination. Site visits suggest that some local programs operate not as a organized system in which classes are scheduled well in advance to allow adequate time for volunteer and participant recruitment, lists of potential (backup) volunteers maintained in the case of volunteer cancellations, and grocery lists for the week planned in advance.

Nevertheless, the quality of OFL partner program staff is one of the greatest strengths of OFL. Staff are extremely hard working, dedicated, flexible, passionate, and energetic. Although most local partners struggle with limited financial, human, and infrastructure resources, staff extend beyond their means to reach intended audiences, offer a multitude of OFL courses, and often fulfill a multiplicity of roles, regardless of the time of day or other obligations. Many local partner staff members have developed effective strategies to leverage resources with organizations for fundraising, both monetary and food donations.

Local OFL programs utilize a wide range of volunteers. In fact, OFL curricula are ideally taught by two volunteers, either two chefs or a chef and a nutritionist. Cooking and food safety topics are not explicitly covered in the *Implementation Guidelines*, because it is assumed that there will be at least one culinary professional per class.

In classes where professional chefs were present, the atmosphere was very engaging and dynamic. These chefs were successful in capturing the attention and interest of participants, regardless of age. Participants interviewed as part of the site visits revealed that they felt special that a professional chef was taking time to work with them and, therefore, had greater interest in learning cooking skills. Without exception, chefs were extremely knowledgeable about cooking and adept at helping participants gain confidence in their own ability to prepare meals at home. Professional chefs were much better than amateur chefs at providing suggestions about low cost strategies. They also encouraged participants to improvise and experiment at home with less expensive or readily available foods. Additionally, professional chefs were highly skilled in demonstrating food safety procedures. They ensured that participants' hands were washed, that work stations were properly disinfected, and that proper knife safety skills were practiced.

However, culinary professionals are not always present. Indeed, professional chefs were present in less than 40% of observed classes and, according to survey data, 25% of staff and 47% of administrators reported that recruiting professional chefs was very problematic. Moreover, many administrators indicated that professional chefs were never present (7%), rarely present (13%), or present only in half the classes (47%), which was largely supported by responses from staff and AmeriCorps volunteers. Yet, without the presence of professional chefs, the quality of OFL implementation is greatly diminished. During site visits, it was evident that some program partners allow the use of "home" chefs, or chefs with little to no professional experience or formal education. In those cases, the importance of cooking and food safety as part of OFL's curriculum were undermined and the chef role reduced to mere execution of recipes. Although survey responses indicated that food safety was adequately taught in most cases, this was not supported in the observation of classes without professional chefs present. By far the most pressing concern is knife safety. In observed classes taught by professional chefs, excellent knife safety skills were taught even though each chef presented the skills somewhat differently.

Corresponding to food and kitchen safety, food budgeting was observed less in visited classes, and survey respondents agreed that this topic is not adequately covered. Observations suggested that professional chefs and nutritionists were superior in incorporating food budgeting into their sessions by providing participants with ideas for alternative

ingredients as well as tips and tricks as to where to locate inexpensive ingredients.

Although OFL's *Implementation Guidelines* do not specify who should teach the nutrition education portion of the curricula, many OFL sites prefer and attempt to use professional nutritionists or dietitians. If professionals are not available, sites try to secure nutrition students from local institutes or universities. These volunteers are much more effective at teaching the nutrition education portion than volunteers who do not possess formal training. They are very confident and can easily manage questions from participants regarding the curriculum. Additionally, they are able to expand upon the OFL curriculum and help participants make decisions about healthy, low-cost food choices. Most program partners prefer to fill the nutrition education role with either professional nutritionists or individuals enrolled in nutrition/dietary programs. However, if program partners are unable to find professional nutritionists or dietitians, they either fill the role themselves or find someone else to do so. In several classes observed during site visits, individuals delivering the nutrition education component without nutrition education backgrounds were not as competent to deliver the OFL curriculum. For example, in one class visited, the instructor did not know how to classify common foods in the associated food groups. While that was the most egregious example, erroneous information was given to participants on numerous occasions although core program staff were present. Even though the nutritional component of OFL does not contain sophisticated concepts, they cannot be adequately conveyed by unqualified individuals.

Discussion

Treatment implementation is not just one thing but rather is a multifaceted process that includes treatment delivery, treatment receipt, and treatment adherence (Cordray & Pion, 2006; Shadish, Cook, & Campbell, 2002). As such, local variations in implementation and service delivery of interventions are an inevitable given the sometimes striking differences in local context, needs, and resources. This is particularly salient for programs and interventions that have been in existence for long periods of time and which are not tightly controlled. Even though some have made a distinction for appropriate evaluation strategies based on program lifecycles (Rossi, Lipsey, & Freeman, 2004), such distinctions are

hard to maintain and justify given that stakeholders' questions of interest cannot always be placed within a lifecycle framework. Well-established programs and interventions are always improvable, even if such improvements are incremental (Weiss, 1998). A crucial strategy for improving such interventions is critically examining assumptions about design and implementation. Even so, receipt of poor or inconsistent program services by participants, for example, can have a large impact on outcomes. Therefore, evaluation of program processes is crucial, but program circumstances typically permit little control over many variables that potentially have substantial influence on outcomes.

Even though a small number of studies have investigated implementation and delivery of nutrition interventions (Davis, Baranowski, Resnicow, et al., 2000; Reynolds, Franklin, Leviton, et al., 2000; Taylor, Serrano, & Anderson, 2001), most prior evaluations and studies of OFL, as well as a number of other nutrition interventions such as Expanded Food and Nutrition Education Program (Burney & Haughton, 2002; Dollahite, Kenkel, & Scott Thompson, 2008), Cooking with a Chef (Condrasky, Graham, & Kamp, 2006), and Eat Smart New York (Colosi, 2007), have mainly emphasized short-term and intermediate outcomes. Largely these studies have disregarded treatment delivery, treatment receipt, and treatment adherence (i.e., fidelity/integrity) and their potential impact on nutritional behaviors and health outcomes and, therefore, have been of limited use for actionable, improvement-oriented decision making. Perhaps most concerning is the use of self-reports (Rohs, Langone, & Coleman, 2001) and retrospective pretests (Reynolds, Franklin, & Leviton, 2000) to measure outcomes, despite the inflationary biases often associated with such methods (Taylor, Russ-Eft, & Taylor, 2009).

Uses of the Evaluation

OFL made extensive use of the evaluation and applied and integrated the study's findings and recommendations into program planning and future program implementation. As a result of the evaluation, OFL increased their national volunteer recruitment efforts for professional chefs, enhanced training and materials around food budgeting, and integrated study tools such as the observation guide into ongoing local and national monitoring efforts.

Limitations

As with all such studies, extrapolating results obtained from small, nonrandomly selected samples to a larger population of interest is extremely difficult, with such difficulties largely due to sampling error. Ideally, a representative sample of sites and classes is one that has strong external validity in relationship to the target population the sample is intended to represent. In the present study, this problem applies mainly to observations of OFL sites and nutrition education classes rather than to questionnaire and extant data, which were predominately census rather than sample data. Therefore, the degree to which observational findings can be generalized with confidence to the population of interest is uncertain. That being said, the biases produced by the inability to randomly select a representative sample of sites and classes are not severe enough to threaten the value of the study's findings or conclusions.

Another relevant limitation to the study is that the instrumentation and measures used to gather data were study-specific. Standardized instruments with known psychometric properties and that also were suited to the nature of the study were unavailable. Even so, interrater reliability and generalizability coefficients for observational measures were well above acceptable limits (Davey, Gugu, & Coryn, 2010, and constructs of interest measured by questionnaire instruments displayed satisfactory levels of internal consistency.

Implications for Research and Practice

Few published examples of process evaluations are to be found in the scholarly literature. This dearth can be attributed, in part, to the view that treatment implementation and delivery is less important than knowledge regarding the outcomes of applied interventions. Outcome studies, in isolation, however, rarely provide adequate information for improving the effectiveness of such interventions. One means for improving outcomes is knowledge about treatment implementation and delivery. Such improvements could, for instance, include modifications to curricula, instructional strategies, or resources and training provided to local implementing partners, among many others. This study demonstrates how program processes, in particular, treatment integrity, can be investigated using mixed methods approaches, integrating multiple types and sources of data. Future research, including comprehensive

process and outcome evaluations, is necessary to assess the degree to which treatment integrity mediates or moderates behavioral outcomes. Such studies would provide valuable empirically-derived knowledge about associations between variances in treatment integrity and the effectiveness of applied nutrition education interventions.

References

- Brandon, P. R. (1998). Stakeholder participation for the purpose of helping ensure evaluation validity: bridging the gap between collaborative and non-collaborative evaluations. *American Journal of Evaluation, 19*, 325-337.
- Burney, J., & Haughton, B. (2002). EFNEP: A nutrition education program that demonstrates cost-benefit. *Journal of the American Dietetic Association, 102*, 39-45.
- Colosi, L. A. (2007). *Eat Smart New York 2007 statewide evaluation report*. Ithaca, NY: Cornell University, Department of Policy Analysis and Management.
- Condrasky, M., Graham, K., & Kamp, J. (2006). Cooking with a Chef: an innovative program to improve mealtime practices and eating behaviors of caregivers of preschool children. *Journal of Nutrition Education and Behavior, 38*, 324-325.
- Cordray, D. S., & Pion, G. M. (2006). Treatment strength and integrity: Models and methods. In R. R. Bootzin & P. E. McKnight (Eds.), *Strengthening research methodology: Psychological measurement and evaluation* (pp. 103-124). Washington, DC: American Psychological Association.
- Coryn, C. L. S. (2007). *Evaluation of researchers and their research: toward making the implicit explicit*. Ph.D. dissertation, Western Michigan University, Kalamazoo.
- Coryn, C. L. S., Gugiu, P. C., Davidson, E. J., & Schröter, D. C. (2008). Assessing needs in hidden populations using respondent-driven sampling. *Evaluation Journal of Australasia, 7*, 3-11.
- Coryn, C. L. S., Noakes, L. A., Westine, C. D., & Schröter, D. C., (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation, 32*(2), 199-226.
- Coryn, C. L. S., Schröter, D. C., & Hanssen, C. E. (2009). Adding a time-series design element to the Success Case Method to improve methodological rigor: an application for nonprofit program evaluation. *American Journal of Evaluation, 30*, 80-92.
- Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation, 80*, 5-23.
- Cullen, A. E. (2009). *The politics and consequences of participation in international development evaluation*. Ph.D. dissertation, Western Michigan University, Kalamazoo.
- Cullen, A. E., Coryn, C. L. S., & Rugh, J. (2011). The politics and consequences of including stakeholders in international development evaluation. *American Journal of Evaluation, 32*(3), 345-361.
- Davey, J. W., Gugiu, P. C., & Coryn, C. L. S. (2010). Quantitative methods for estimating the reliability of qualitative data. *Journal of MultiDisciplinary Evaluation, 6*(13), 140-162.
- Davidson, E. J. (2005). *Evaluation methodology basics: The nuts and bolts of sound evaluation*. Thousand Oaks, CA: Sage.
- Davis, M., Baranowski, T., Resnicow, K., et al. (2000). Gimme 5 fruit and vegetables for fun and health: process evaluation. *Health Education & Behavior, 27*, 167-176.
- Dollahite, J., Kenkel, D., & Scott Thompson, C. (2008). An economic evaluation of the Expanded Food and Nutrition Education Program. *Journal of Nutrition Education and Behavior, 40*, 134-143.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed). Thousand Oaks, CA: Sage.
- Reynolds, K. D., Franklin, F. A., Leviton, L. C., et al. (2000). Methods, results, and lessons learned from process evaluation of the High 5 school-based nutrition intervention. *Health Education & Behaviour, 27*, 177-186.
- Rohs, F. R., Langone, C. A., & Coleman, R. K. (2001). Response shift bias: a problem in evaluating nutrition training using self-report measures. *Journal of Nutrition Education, 33*, 264-292.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed). Thousand Oaks, CA: Sage.
- Schröter, D. C. (2008). *Sustainability evaluation: development and validation of an evaluation checklist*. Ph.D. dissertation, Western Michigan University, Kalamazoo.
- Scriven, M. (2007). *Key evaluation checklist*. Kalamazoo, MI: The Evaluation Center.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

- Share Our Strengths Operation Frontline. (2009). *Implementation guidelines*. Washington, DC: Share Our Strengths.
- Swindle, S., Baker, S. S., & Auld, G. W. (2007). Operation Frontline: assessment of longer-term curriculum effectiveness, evaluation strategies, and follow-up methods. *Journal of Nutrition Education and Behavior*, 39, 205-213.
- Taylor, P. J., Russ-Eft, D. F., & Taylor, H. (2009). Gilding the outcome by tarnishing the past: inflationary biases in retrospective pretests. *American Journal of Evaluation*, 30, 31-43.
- Taylor, T., Serrano, E., & Anderson, J. (2001). Management issues related to effectively implementing a nutrition education program using peer educators. *Journal of Nutrition Education*, 33, 264-292.
- Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.