

Evaluation of a USDA Nutrition Education Program for Low-income Youth

Marilyn S. Townsend, PhD, RD¹; Margaret Johns, MPA, RD²;
Mical Kay Shilts, PhD³; Lucrecia Farfan-Ramirez, MPH⁴

ABSTRACT

Objective: Examine effectiveness of a state's Youth Expanded Food and Nutrition Education Program (EFNEP) and assess the validity of the federal impact indicator method for reporting program outcomes.

Design: A randomized, controlled field trial of 229 groups with 5,111 youth, 9-12 years old, in community settings.

Intervention: 6- to 8- hour, 7-lesson education experience with food preparation and tasting, an education experience typical of EFNEP in California.

Outcome Measures: US Department of Agriculture (USDA) impact indicators: nutrition knowledge, eating a variety of foods, food selection, and food preparation and safety practices.

Analysis: Analysis of covariance model controlling for pretest, gender, age, and ethnicity, with group nested in condition.

Results: Organizing responses by impact indicators, treatment participants made significant gains on the posttest compared to controls for 3 of 4 indicators ($P < .008$ to $P < .0001$). Gains were made by 34 to 68% of youth participants for 4 indicators. The impact indicator method for federal reporting compared favorably with results from a randomized controlled trial with groups nested in conditions.

Conclusion and Implications: This is the first report in the literature of (1) a large evaluation study of Youth EFNEP and (2) an estimate of the validity of the USDA impact indicator method for reporting program outcomes.

Key Words: EFNEP, evaluation, youth, children, impact indicator, USDA

(*J Nutr Educ Behav.* 2006;38:30-41.)

INTRODUCTION

The Expanded Food and Nutrition Education Program (EFNEP) is a primary prevention, health promotion intervention under the guidance of the United States Department of Agriculture (USDA),¹ targeting low-income families and youth in 50 states and 6 US territories. Its major objective is to "assist these adults and youth in acquiring knowledge, skills, and behaviors necessary for nutritionally sound diets, contribute to their personal development and

the improvement of the total family diet and nutritional well-being." Approximately \$62 million is appropriated to EFNEP each year in the Farm Bill legislation. With an experiential learning approach, the youth component of EFNEP (ie, Youth EFNEP) is delivered to children ages 3 to 19 years as a series of lessons over a period of days, weeks, or months at sites that include schools, after-school care, youth clubs, day camps, residential camps, community centers, and home/community gardens. In fiscal year 2001, over 447 000 youth in 16 700 groups participated in Youth EFNEP.¹

State EFNEP programs are evaluated annually by USDA using USDA-selected impact indicators. Data for all states and territories are combined by the program leader in the Office of Research, Education and Economics, USDA, and submitted annually to Congress in the form of a formal report.¹

In a major review of nutrition education program evaluations, Contento et al reported on 43 studies for programs

¹Department of Nutrition, University of California, Davis

²University of California Cooperative Extension, Kern County, CA

³Family and Consumer Science Dept., California State University, Sacramento, CA

⁴University of California Cooperative Extension, Alameda County, CA

Address for Correspondence: Marilyn Townsend, PhD, RD, UCD Nutrition Dept., 1 Shields Ave., University of California, Davis, CA 95616-8783; Phone: (530)754-9222; Fax: (530) 752-7588; E-mail: mstownsend@ucdavis.edu

©2006 SOCIETY FOR NUTRITION EDUCATION

doi: 10.1016/j.jneb.2005.11.014

involving elementary school-aged children.² Of these, 8 were outcome evaluation studies of school-based general nutrition education programs that overlapped with the 9- to 11-year-old age group.² None of these studies was for Youth EFNEP. Hoelscher identified 10 evaluation studies for 9- to 11-year-olds, and again, none was for Youth EFNEP.³ Our review of the literature, using the key words *EFNEP* and *evaluation* with 6 databases, identified no evaluation studies of Youth EFNEP. We were not surprised by the lack of published studies, because no separate source of funds is allocated at the USDA level for a more sophisticated research design for program evaluation. Although the research design for the annual evaluation (ie, pre/post, 1-group design) satisfies the requirement for the USDA-mandated reporting process, it does not meet the standard for a study to be published in a peer-reviewed journal such as the *Journal of Nutrition Education and Behavior*. A more sophisticated research design for evaluation to meet the standards for journal peer review would require^{4,5,6,7}:

- (1) Random sampling of all program participants, or equivalent;
- (2) Treatment and control conditions;
- (3) Random assignment to condition; and
- (4) A sufficiently large number of groups of participants for the hypothesis tests to have adequate power when cluster effects exist.

Although USDA legislation for EFNEP is dated 1969, our study is the first report of an evaluation of a statewide EFNEP program for 9- to 11-year-old youth. Our primary purpose was to examine program effectiveness of California's Youth EFNEP as it is normally delivered in the community, using a research design that incorporates the 4 factors presented above and that meets the rigor of a peer-reviewed journal. Our secondary purpose was to estimate the validity of the widely used USDA impact indicator method for program evaluation—a task that has not been undertaken by any previously published study. The USDA impact indicator method for program evaluation is used across all land grant university Cooperative Extension programs, not just EFNEP. Therefore, our study will be the first to report evaluation findings for a youth EFNEP study using a research-quality protocol and the first to estimate the validity of the USDA impact indicator method for reporting program outcomes. This study is important for continued funding for youth EFNEP and continued use of the USDA impact indicator method.

METHODS

The 10 counties in California receiving Youth EFNEP funds participated in the evaluation study from October 1998 to October 1999. The target audience was low-income children, aged 9 to 11 years, participating in EFNEP. The program delivery model included training of leaders by EFNEP staff, followed by leader delivery of the intervention

to youth participants. As is the usual procedure in California, leaders were recruited by Cooperative Extension staff at schools, community centers, and daycare programs, among other community locations. Leaders included classroom teachers, after-school program staff, summer day camp staff, community agency personnel, and select teenagers. Use of the term *group* refers to a collection of children participating in the study. Groups were located in schools, at summer and inter-session day camps, and as part of community after-school programs. All leaders and youth in the 9-11-year-old age category who were enrolled in EFNEP in California for FY 1998-1999 participated in the study. Random selection of participants was not deemed applicable, because all youth enrolled in EFNEP in this age category participated in the study.

Sample size

Our goal was to detect a difference in pre/post scores of 10 percentage points with 80% power and 5% level of significance. Sample size calculations generated from data collected in 1996 and 1997 determined that for the number of youth groups, 200 treatment and control youth groups would be excellent, while 100 would be marginal and 50 inadequate.⁸ Of the 300 youth groups targeted for recruitment, 200 would be randomized to treatment and 100 to control. The 2:1 ratio of treatment to control groups allowed us to analyze results by race and age of participants, without increasing the number of groups in the study.

Research design

This evaluation study used a randomized, controlled research design, with the youth group serving as the unit of assignment, intervention delivery, and analysis.^{3,4,5,6} Upon successful recruitment of a group leader, the field staff person notified the research staff at the EFNEP state office in the Nutrition Department at the University of California, Davis. Random assignment was determined using the random numbers table. Leaders and participants were recruited over a 12-month period. To assist with leader compliance to our research protocol, \$25 certificates for a local supermarket chain were provided to participating leaders with instructions to use the funds for food supplies for the nutrition activities (eg, fresh produce). Leaders assigned to the treatment condition were then trained to use the curriculum components of the education intervention. Early in the study, we found that leaders generally did not like to be randomized to the control condition. Consequently, we instructed EFNEP staff to use a new term, "delayed intervention," instead of "control."

Intervention description

The education intervention titled *Eating Right Is Basic for 9-11 year olds* (ERIB) was adapted from the national EFNEP

publication, *Eating Right is Basic* (Michigan State University Extension), for use in California and contained seven lessons to be taught in approximately 6 to 8 weeks.⁹ This education experience is typical of EFNEP in many other states.¹ The intervention focused on increasing awareness of fruits and vegetables in healthful diets and of food safety practices for children who prepare some of their own meals and snacks. The intervention was an example of a knowledge curriculum defined by Contento as: "Those directed at general nutrition education and based on the paradigm that new knowledge leads to attitude change which, in turn, leads to behavior change."^{2,10} The curriculum encompasses lessons to enhance knowledge, skills, and food choices via experiential activities that include food tasting, food art, food puzzles, games, and preparation of fruit and vegetables. To recruit leaders and elicit a commitment, the study was explained thoroughly to prospective leaders. As is the customary procedure for Youth EFNEP, field staff conducted 1- to 2-hour workshops for these committed leaders using a train-the-trainer model with a specific curriculum designed for this training.¹¹ At the training, staff modeled the experiential techniques that they were recommending for use with participants. The study purpose, design, and random assignment were described again, with emphasis on maintaining study fidelity. As recommended by Dennis, incentives were provided to increase the likelihood of complete intervention implementation.¹²

Instrument development and testing

Nutrition and food safety knowledge and food preparation skills were assessed by self-report with the evaluation instrument *Kids Kartoons*, a cartoon-style booklet designed specifically for this education intervention.¹³ Test items were constructed by the investigators so that instructional content from each lesson was represented on the test, important instructional topics were well represented, and a balance was achieved among the broader concepts. Multiple choice items were selected as the preferred response option format to reduce respondent burden during test administration.

Two items were placed on each page of the 8-page, reusable booklet. Participants indicated their answers on a 3-page, scannable response sheet. In an analysis of the readability of the survey, a Flesch Kincaid score of 2.8 and a Flesch Reading Ease score of 91.5 indicated that a participant reading level of 3rd grade was required for understanding survey items.^{14,15} Since this survey was not intended to be an assessment of reading skills, EFNEP staff read the items aloud in many cases and circulated among participants to offer assistance with the understanding of each item.

Content and face validity. Content or representative validity exists when an evaluation instrument provides suitable coverage of the universe of knowledge, skills, be-

haviors, etc., that the test is designed to measure.^{16,17} To assess content validity, 9 nutrition experts with varied backgrounds and familiar with this education intervention reviewed the initial instrument to verify that items reflected the content of the intervention.¹⁶ Of the experts, 3 were members of a university nutrition department, 3 were Cooperative Extension academic advisors familiar with youth programming responsibilities, and 3 were Cooperative Extension field staff familiar with leaders and youth recruited for the program. Each expert verbally provided his opinion about content to the first author, and the recommended changes were made. Face validity implies that the evaluation instrument appears practical, pertinent, and relevant to the purpose of the test and is assessed by members of the target audience.^{16,17} Using individual interviews with one of the authors, 8 youth from the target audience were interviewed about the questionnaire. In response to the youths' verbal comments, some changes to wording were made in order to reduce ambiguity and maximize the clarity of the questions.¹⁸ Copies of the questionnaire booklet are available from the first author.

Field testing and item analysis. The questionnaire was field tested with over 750 youth in 1996 and retested with another 700 youth in 1997.¹⁹ Item analysis generating an item difficulty index indicates the extent to which participants answer an item in the same way.²⁰ Items are not considered useful if they are answered correctly by more than 80% (item difficulty index = 0.80) or fewer than 20% (item difficulty index = 0.20) of participants.²¹ Consequently, items were retained with mean values between 0.2 and 0.8 on a range of 0 to 1.0. Each year two items were eliminated or reworded and retested the following year.

Reliability. The assessment of reliability was intended to establish that the evaluation tool was measuring phenomena in a reproducible and consistent way.^{16,17} Using the test-retest approach to reliability, we assessed the degree to which the test scores were free from random error.²² Using the method described by Carmines, a group of participants ($n = 120$) sharing the same characteristics as those in the main study completed the survey on 2 occasions, 3 weeks apart, with no intervention.¹⁷ Eighty-seven matched pre and posttests were available for analysis.

Interpretation of the reliability score depends on the range of ability of participants in the sample, sample size, and intended use.^{17,21} When using an instrument for evaluation with groups of children, the instrument is sufficiently sensitive and reliable at the 0.6 level.¹⁷ Although merely adequate, the stability of the instrument was sufficient for its intended use, with an overall Pearson Product Moment correlation coefficient of 0.62.

Survey administration

To ensure consistency of data collection procedures across youth groups, EFNEP field staff were trained to administer

the pre and post survey to the participants using a protocol designed for this study.²³ Staff responded to all questions from participants. This protocol was followed with all groups, with the exception of 2, where the location was more than 50 miles away. In those 2 instances, the leader administered the post survey. Following the pretest survey, the leaders in the treatment condition were instructed to proceed with the intervention during the next 6 to 8 weeks. Each lesson contained at least 1 hour of content and could be delivered in segments, or all at once, depending on the leader's schedule. If time permitted, the leader could choose to expand the lesson with additional suggested activities. Six weeks later, EFNEP field staff also administered the posttest survey to participants, and the implementation questionnaire to leaders. Leaders in the "delayed intervention" control condition were instructed not to teach the intervention until after the posttest survey was administered.

Intervention implementation

Monitoring the implementation of our education intervention should enhance interpretation of the outcome data.²⁴ At risk is evaluation of an intervention that is inadequately implemented. Failure to assess implementation can lead to what has been termed type III error, where weak or null results are attributed to intervention failure, when in reality the intervention was improperly implemented.²⁵ Most implementation tools focus on two dimensions: quantity (dose)—how much of the actual intervention was delivered to participants; and quality (fidelity)—to what extent the intervention was delivered as intended.²⁴ To document *dose*, we asked leaders to identify: (1) the number of lessons delivered, of a possible 7; (2) the number of activities completed, of a possible 17; (3) total number of hours of instruction; and (4) time devoted to lesson preparation. To document *fidelity*, we asked leaders to identify the curriculum components implemented: (1) use of the series of parent letters designed to involve parents and caregivers in their children's activities; (2) implementation of food preparation activities; and (3) implementation of food tasting experiences. In all cases, the data was self-report and collected retrospectively after completion of the program.

Analysis for evaluation study

To facilitate the accuracy and cost of data entry, the answer sheets for the evaluation instrument were created in TELEform, version 1.0 (Cardiff Software, Inc., Vista, CA).²⁶ TELEform allows the answer sheet to be scanned and data to be automatically entered into the data tables in Microsoft Access, version 6.0. With approximately 31 000 data sheets (6 per participant) involved in the study, data entry time was greatly reduced using the TELEform system. Once in Microsoft Access, most data were analyzed via SAS, Version 8.2 (SAS Institute, Inc., Cary, NC). Analysis

for reliability generated a Pearson Product Moment correlation coefficient using SPSS for Windows, Version 10.0 (SPSS, Inc., Chicago, IL). Statistical significance for all analyses was $P \leq .05$.

In terms of an overall estimate of intervention effect, *group* was the most appropriate unit of analysis, given that groups, not individual participants, were randomized, and all aspects of the intervention were delivered with groups. Consequently, we randomized groups of children and conducted the intervention with groups. To incorporate individual-level variables, participant data were included in the model, with *group* included as a random effect, as recommended by Murray.⁵ Group was nested within the condition of treatment or control. As a result, the denominator degrees of freedom were based on the number of groups, not the number of participants, and the actual denominator itself reflected the between-group variability, not the between-child variability.⁶

Since randomization at the group level made it difficult to account for differences between groups without using a large number of groups, our sample size calculations took into account this design and involved calculating the number of groups needed, as well as the number of children needed in each group.^{5,6} To account for between-group variation, emphasis was placed on additional groups, instead of additional participants, when field staff had the option.

Because there was no evidence that participants might not be similar on all uncontrolled variables, pretest scores were used in an analysis of covariance to adjust group means.²⁷ The explanatory variable was condition ("intervention" treatment or "delayed intervention" control) as main effect with covariates being pre-intervention score, gender, age, and ethnicity, with group nested in condition (SAS Institute, Inc., SAS/STAT User's Guide, Version 8, Cary NC: SAS Institute Inc., 1999). The response variable was postintervention scores in all analyses. The total sample was stratified by sex and ethnicity/race, and the analyses were repeated for males, females, and 3 major ethnic/racial categories using the same model. Procedures of the Institutional Review Board for University of California, Davis were followed for this study.

RESULTS

Sample characteristics

Completing pre and post surveys for the study were 246 groups containing a total of 5508 children. Although only treatment leaders were asked to complete the Leader Implementation Questionnaire, our analysis revealed that 17 leaders assigned to the "delayed intervention" control condition also completed the survey and indicated that they introduced some or all of the intervention to participants during the "delayed intervention" period. Consequently, the data from those 17 youth groups were removed, so that

the final analyses were conducted with 229 youth groups of 5111 participants (n = 229). Results are presented for 162 intervention groups, with 3586 children (n = 162), and 67 delayed intervention control groups, with 1526 children (n = 67), from the 10 counties with EFNEP programming.

Almost 51% of participants were female and 94% were between the ages of 9 and 11 years, with an average age of 10.0 ± 1.0 years (Table 1). The largest ethnic category of participants was Hispanic (43%).

USDA impact indicator method for program evaluation

USDA guidelines require that evaluation data be collected and reported to the federal government annually for one or more of the four USDA impact indicators: (1) eat a variety of foods as a result of the program; (2) show increased knowledge of human nutrition; (3) show increased ability to select low-cost, nutritious foods; and (4) show improved knowledge and practices in food preparation and safety. Survey items were grouped by impact indicators (Table 2).

Participant responses were totaled for pre and posttests by each impact indicator. Data were then analyzed for children in the treatment condition (n = 3586) using the USDA method where we calculated the percentage of children showing higher posttest than pretest scores (Table 4). Specifically, there were favorable gains for each indicator: 34% of children had improved scores for *Eat a Variety of Foods*, 53% for *Nutrition Knowledge*, 31% for *Food Selection*, and 68% for *Food Preparation Skills and Safety Practices*.

Table 1. Demographic Characteristics of Participants in Youth EFNEP Evaluation Study (n = 5,111).

Characteristic	n	%
Gender		
Male	2521	49.3
Female	2590	50.7
Race/ethnicity		
Non-Hispanic black	664	13.0
Non-Hispanic white	961	18.8
Hispanic	2193	42.9
Asian	614	12.0
Native American	153	3.0
Other	526	10.3
Age (y)*		
7	7	0.1
8	142	2.7
9	1426	30.8
10	1913	37.4
11	1325	25.9
12	291	5.6
13	5	0.1
14	1	0.0

*Data for age missing for 1 participant.

Other child outcomes

While health promotion intervention efficacy studies are designed to evaluate what an intervention achieves under optimum conditions, effectiveness studies, such as this study, are concerned with testing whether the intervention does more good than harm when delivered via a real-world program.²⁸ To determine the latter, we compared treatment to control participants for the four USDA impact indicators using analysis of covariance (ANCOVA) controlling for covariates and group nested in condition (Tables 3 and 4). Participants receiving the intervention produced greater gains compared to control participants for 3 of the 4 indicators (P < .008 to P < .0001) (Table 4).

We further examined these outcomes for each indicator by age, ethnicity, and gender. Females scored significantly higher than males on the posttest using total scores (P < .0001) (not shown). Statistically significant gains were found for females for 2 indicators and males for 3 indicators (Table 3). Race/ethnicity also had a significant impact on outcome, with white participants scoring the highest on the posttest using total scores (P < .0001) (not shown). Statistically significant gains (P < .0001) for 2 indicators were found for non-Hispanic whites (n = 186 groups) and Hispanics (n = 213 groups) (Table 3). For non-Hispanic blacks, 1 indicator was marginally significant (n = 161 groups, P = .06). Age did not have a significant impact on outcome (P = .31) (not shown).

Intervention implementation

Leaders in the treatment condition were asked to document the activities they conducted with their participants. Of those leaders, 80% (n = 128) completed the questionnaire. Use of components of the education intervention was reported by leaders to determine effects of those components on child outcomes. Although a positive trend was noted with an increased number of activities and posttest scores (12 activities vs. 13 activities), the relationship was not statistically significant (P = .09). No difference in scores was found for use of the parent letter series (P = .37), nor for number of hours of instruction (P = .18). Surprisingly, youth groups participating in food preparation experiences received lower scores on the posttest (P = .03). The leader implementation scores were not predictive of outcomes, as indicated by children’s posttest scores.

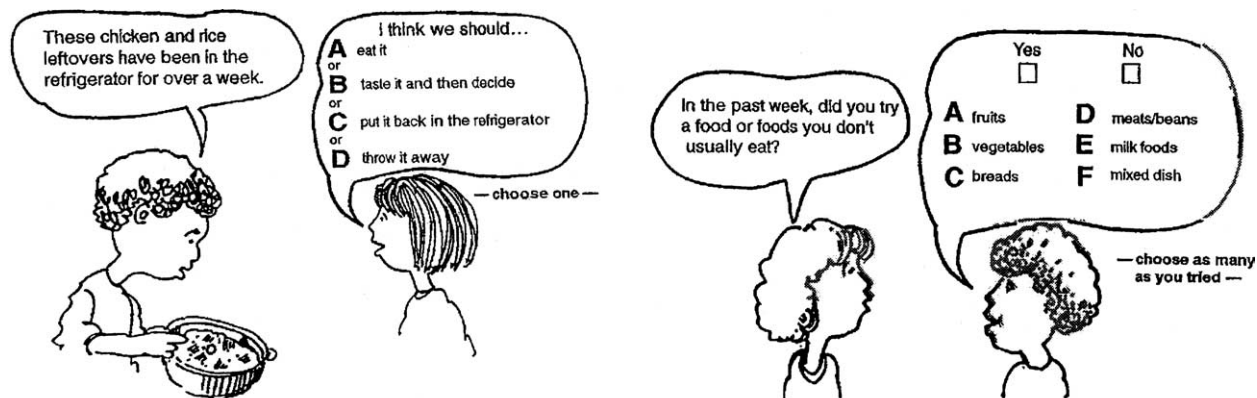
Fidelity to the study design. Evidence from the Leader Implementation Questionnaire suggested that at least 17 “delayed intervention” control leaders taught some or all of the lessons during the period when they were to refrain. These control leaders identified specific activities conducted during the delay period prior to data collection, resulting in contamination of the control. Where evidence was strong from the Leader Implementation Questionnaires, we removed those groups. However, we made a

Table 2. Evaluation Survey Items, Response Options, and Points Grouped by USDA Impact Indicators.

USDA Impact Indicator	Item	Maximum score (Points)
Eat a variety of foods	I don't know what it means to eat a variety of foods. I think it means. . . .	7
	<ul style="list-style-type: none"> ● you should eat the same nutritious foods at every meal; ● you should eat many different types of foods. 	1
	In the past week, did you try a food or foods you don't usually eat? . . . Yes, No. What were they? Choose as many as you tried . . .	6
	<ul style="list-style-type: none"> ● Fruits ● Vegetables ● Breads ● Meats/beans ● Milk foods ● Mixed dish 	
Nutrition knowledge	I don't know what it means to eat a variety of foods. I think it means. . . .	6
	<ul style="list-style-type: none"> ● you should eat the same nutritious foods at every meal; ● you should eat many different types of foods. 	1
	I want to make a sandwich with fruit, a vegetable, and a high-fiber food. That's a good idea. We can make sandwiches together. Let's have . . .	1
	<ul style="list-style-type: none"> ● Tuna with some mayonnaise, raisins, lettuce, and whole wheat bread; ● Tuna with some mayonnaise, lettuce, tomato, and white bread; ● Tuna, white bread, and cheese. 	
	I'm skipping breakfast today. But skipping meals is not good for you because . . .	1
	<ul style="list-style-type: none"> ● Your body needs food throughout the day so you don't get tired and grouchy; ● You should eat exactly three meals a day. 	
	I wonder why there are food commercials on T.V. Food companies advertise. . . .	
	<ul style="list-style-type: none"> ● To tell you if the food is good for you; ● To teach people about nutrition; ● So people will buy their products. 	
	My sister wants to eat foods low in fat. I know some foods that are low in fat . . .	1
	<ul style="list-style-type: none"> ● Hamburger, cheese, catsup, and 2% milk; ● Cereal with 2% milk and an apple; ● Cheese pizza and an apple; ● Taco at a fast food place and 2% milk. 	
Which food is not high in fiber? . . .	1	
<ul style="list-style-type: none"> ● Popcorn ● Beans ● Broccoli ● Cheese 		
Food selection	In the past week, did you try a food or foods you don't usually eat? . . . Yes, No. Choose as many as you tried . . .	3
	<ul style="list-style-type: none"> ● Fruits, ● Vegetables, ● Breads, ● Meats/beans, ● Milk foods, ● Mixed Dish 	1
	Choose one.	1
	<ul style="list-style-type: none"> ● I planned, prepared, and tasted a nutritious snack this week; or ● I did not have time to plan, prepare, and taste a nutritious snack this week. 	
	I ate breakfast or a snack before school today . . .	1
<ul style="list-style-type: none"> ● Yes ● No 		
Food preparation skills & safety practices		10
	What is the best way to turn the handle of a pan when it's on the stove? . . .	1
	<ul style="list-style-type: none"> ● Away from the edge of the stove; ● Toward the edge of the stove; it doesn't matter. 	

(continued)

USDA Impact Indicator	Item	Maximum score (Points)
	I'm cutting an apple. What is the best way to hold a knife? I cut with the knife. . . . <ul style="list-style-type: none"> ● Away from me; ● Toward me on a cutting board; ● Away from me on a cutting board. 	1
	My dad left these foods for me to put away. Please help me. <i>Leftover pizza</i> <ul style="list-style-type: none"> ● Counter; ● Refrigerator; ● Freezer. 	1
	My dad left these foods for me to put away. Please help me. <i>Yogurt, unopened</i> . . . <ul style="list-style-type: none"> ● Cabinet; ● Refrigerator; ● Freezer. 	1
	My dad left these foods for me to put away. Please help me. <i>Cooked rice, covered</i> . . . <ul style="list-style-type: none"> ● Counter; ● Refrigerator; ● Freezer. 	1
	My dad left these foods for me to put away. Please help me. <i>Potato, raw</i> . . . <ul style="list-style-type: none"> ● Cabinet; ● Refrigerator; ● Freezer. 	1
	I'm going to school now. By the way, what are we having for dinner tonight? Chicken, but it's frozen, so I am going to let it thaw . . . <ul style="list-style-type: none"> ● On the counter; ● In the refrigerator; ● In the sink. 	1
	These chicken and rice leftovers have been in the refrigerator for over a week. I think we should. . . . <ul style="list-style-type: none"> ● Eat it; ● Taste it and then decide; ● Put it back in the refrigerator; ● Throw it away. 	1
	My mom gave me these kitchen utensils. Match the utensil with the picture. <ul style="list-style-type: none"> ● Grater ● Vegetable peeler ● Wok 	1
	I ate a lot of food at the picnic and now I have a stomachache and feel awful . . . The food was left out of the cooler all day. I'll bet you have . . . <ul style="list-style-type: none"> ● The flu; ● Food poisoning; ● Nerves because of the math test tomorrow. 	1



Examples of cartoon style format of 2 survey items with response options.

major tactical error in not asking all leaders, including those assigned to be controls, to complete the Leader Implementation Questionnaire. Consequently, for the 67 con-

trol leaders remaining in the analyses, we do not know with certainty whether they complied with their “delayed intervention” assignment.

Table 3. Measuring Impact for 4 USDA Indicators for 2 Sex and 3 Ethnic/Racial Categories.*†

Impact Indicator	Change in Treatment‡	Change in Control‡	p
Eat a Variety of Foods			
Males	0.07 ± 1.54	0.11 ± 1.39	0.58
Females	0.16 ± 1.45	0.22 ± 1.32	0.25
Non-Hispanic Black	0.04 ± 1.44	0.21 ± 1.31	0.97
Non-Hispanic White	0.07 ± 1.51	0.19 ± 1.42	0.67
Hispanic	0.19 ± 1.41	0.16 ± 1.29	0.88
Nutrition Knowledge			
Males	0.67 ± 1.45	0.20 ± 1.24	<0.0001
Females	0.63 ± 1.38	0.30 ± 1.26	<0.0001
Non-Hispanic Black	0.62 ± 1.41	0.28 ± 1.21	0.22
Non-Hispanic White	0.68 ± 1.35	0.26 ± 1.17	<0.0001
Hispanic	0.65 ± 1.44	0.24 ± 1.30	<0.0001
Food Selection			
Males	0.04 ± 1.02	-0.04 ± 1.01	<0.001
Females	0.11 ± 1.0	0.05 ± 0.97	0.94
Non-Hispanic Black	-0.03 ± 0.99	0.09 ± 1.04	0.11
Non-Hispanic White	0.10 ± 0.99	-0.02 ± 0.97	0.82
Hispanic	0.10 ± 1.02	0.002 ± 1.002	0.21
Food Preparation Skills & Safety Practices			
Males	1.02 ± 1.70	0.42 ± 1.59	<0.0001
Females	0.92 ± 1.66	0.38 ± 1.44	<0.0001
Non-Hispanic Black	0.84 ± 1.60	0.40 ± 1.63	0.06
Non-Hispanic White	1.08 ± 1.51	0.56 ± 1.27	<0.0001
Hispanic	0.94 ± 1.71	0.24 ± 1.52	<0.0001

*ANCOVA model controlling for pretest, gender, age, and ethnicity, with group nested in condition. Answers were coded so that a desirable response received a higher score.

†Sample size for male participants in the treatment (1762) and control (759); Female participants in the treatment (1823) and control (767). Sample size for treatment condition for non-Hispanic white (632), non-Hispanic black (449), and Hispanic (1548) youth. Sample size for 'delayed intervention' control condition for non-Hispanic white (329), non-Hispanic black (215), and Hispanic (645) youth.

‡Unadjusted means

Validation of USDA impact indicator method

Validation of this method requires comparing it to a rigorous approach using data from the same EFNEP participants.^{17,21} As our rigorous approach, we used ANCOVA, controlling for covariates and group nested in condition. Treatment participants scored greater gains than the control participants for: (1) nutrition knowledge ($P < .0001$); (2) food selection ($P < .008$); and (3) food preparation and safety practices ($P < .0001$) (Table 4). Although 34% of participants improved their scores for the indicator "Eat a variety of food," an examination of the change in mean scores using the rigorous method indicates that no statistically significant difference occurred ($P = .82$). The rigorous method (ie, ANCOVA) supports the results for 3 of the 4 USDA impact indicators.

DISCUSSION

Youth EFNEP is a primary prevention, health promotion intervention targeting low-income families and youth. We

conducted a study of this USDA program as it is normally delivered in the community in California, as opposed to an efficacy study examining optimum program delivery by paid, trained staff. Treatment and control posttest scores were significantly different in our study, with the treatment groups making greater gains than the control groups for 3 indicators (Table 4). Because of the study design characteristics, we attribute those gains to the intervention. Although small, the gains should be celebrated, given the nature of this brief intervention and its relatively low intensity. In addition, most leaders were teachers in school settings who already had responsibility for many other areas of their students' academic needs. Given our relatively low-intensity intervention, expecting greater gains may be unrealistic.

Means for students' pretest scores were relatively high (12.31 points out of 21 total, or 59% of answers were correct), leaving little room for improvement. In previous years' field testing, when the instrument was tested and refined with 9- to 11-year-old EFNEP members, mean pretest scores were lower, in the range of 30-40 percentage points. Given that the intervention and evaluation instru-

Table 4. Comparison of results from USDA impact indicators (treatment group pre/post, % participants) with treatment/control group pre/post means to assess validity of USDA indicator method.

Impact Indicator	USDA Impact Indicator Method*		ANCOVA Method †		p-value
	Treatment group scores (n = 3586 participants)		Treatment (n = 3586)/control (n = 1525) in 229 groups		
	Participants making ≥ 1 positive change	n (%)	Change in Treatment	Change in Control	
Eat a variety of foods	1243 (34)	0.12 ± 1.50	0.16 ± 1.35	0.82	
Nutrition knowledge	1905 (53)	0.65 ± 1.42	0.25 ± 1.25	<0.0001	
Food selection	1113 (31)	0.07 ± 1.01	0.01 ± 0.99	<0.008	
Food preparation skills & safety practices	2428 (68)	0.97 ± 1.68	0.40 ± 1.52	<0.0001	

*USDA method of reporting impact data: % participants in program who change from pre to post data collection periods.

†ANCOVA method is used as the 'gold standard' to validate the USDA method of reporting impact data: ANCOVA model controlling for pretest, gender, age, and ethnicity, with group nested in condition.

#Unadjusted means.

ments were essentially the same during field testing and the study, we cannot account for the unanticipated higher pretest scores with this year's participants.

Sex and ethnic differences. To explain the difference by sex, female participants may have been more interested in the topics covered by the intervention or more receptive to mass media information. To enhance effectiveness with male participants, effort should be made to include topics that would engage and motivate them as well as the female participants. Qualitative research could shed light on the motivators specific to males. Sample sizes were smaller for the analyses stratified by ethnic/racial group and may have made detecting statistically significant differences very difficult, particularly for the analyses for non-Hispanic black children. Of the 3 ethnic groups, non-Hispanic black children were only in 161 groups, generating less power for the analysis. Our sample size calculations indicated appropriate numbers for treatment and controls for the analysis with the entire sample, but not for the stratified samples.

Comparison to other studies. In the 8 studies of general nutrition education programs identified by Contento et al, the intervention dose varied from 8 lessons in 8 weeks to 24 lessons in 6 months.² The dose for these 8 studies was greater than for the 7 EFNEP lessons in our study. These 8 studies all demonstrated nutrition knowledge gains and, for some programs, modest positive changes in short-term behaviors. Our results are similar, with positive changes in knowledge and short-term behaviors. As with our study, these other studies were also set in a real-world context, where program leaders were volunteers and not paid research staff.

Validation of the USDA impact indicator method. The USDA method of program evaluation is mandated for EFNEP and many other Cooperative Extension programs. Although a secondary objective of our study, providing evidence for the validity of the USDA impact indicator method is an important finding. Our EFNEP study is the first report of such evidence. As is the customary procedure for USDA reporting, we reported percentage (%) of youth making changes for each of 4 impact indicators (Table 4). However, with the application of our more rigorous model (ie, the "gold standard") that included the control condition and analysis of groups nested in condition, the results continued to be statistically significant for 3 of the 4 indicators. A closer examination of the 4th indicator, *Eat a variety of foods*, showed that participants in the control condition made more gains than did children in the treatment condition. With the inclusion of the control group in the analysis, the final model is not statistically significant.

These favorable results suggest that the USDA reporting method is fairly accurate, although with some bias

favoring EFNEP. Given the ease with which the USDA method can be implemented by state programs, our finding is important as an estimate of the validity of this method. The results of our rigorous analysis offer evidence for the continued use of the USDA impact indicators method, with its relative ease of implementation and minimal administrative burden for states.

Implementation and child outcomes. Conducted concurrently with our study, Resnicow et al examined the validity of 3 measures of implementation: classroom observation of fidelity, teacher self-report questionnaire, and postimplementation interview.²⁹ Resnicow found that the teacher self-report questionnaire was not a valid measure of implementation completeness. Likewise, we found that our data from the Leader Implementation Questionnaire was problematic.

Surprisingly, we found that those children participating in food preparation had lower total posttest scores than other participants, and that total scores on the Leader Implementation Questionnaire were not predictive of participant outcomes. The dose-related results are equally difficult to explain. We considered two possible reasons for these disturbing results. First, leaders who conducted food preparation and tasting activities displaced some of the other nutrition lessons with food preparation instead of increasing the number of activities. But our second and most likely reason is that our method of collecting leader implementation data was unreliable. We asked leaders to report two months after the onset of the intervention. A limitation of our method was its retrospective recall of information. We asked leaders 6 to 8 weeks after their training workshop to recall specifics of the educational experience. A recall method is more likely to reflect a social desirability bias from pressure to teach more lessons.²⁹ In hindsight, we should have used a record or log method, where leaders record as they conduct activities, or better yet, conducted in-person leader interviews. Based on Resnicow's work, we surmise that leader interviews with probing questions would be most helpful for collecting more accurate implementation data for determining how to improve program quality of our intervention.²⁹ Because of the poor quality of the implementation data, we cannot estimate the predictive validity of the various components of the intervention.

We focused our energy and resources on analyzing psychometric properties of the child survey instrument and failed to do the same for the Leader Implementation Questionnaire. Clearly, results suggest that data provided by the leader retrospectively is not as valid and reliable as researchers require.

Fidelity to the study design. According to Devine, however tightly researchers attempt to control placement of subjects of community trials in experimental and control conditions, some "leakage" inevitably occurs.³⁰ The task is

to keep this leakage to a minimum to maintain the internal validity of the study.¹² We have evidence that some leaders who were randomized to the "delayed intervention" control were offering program activities during the time of the study. Although agreeing to terms of study participation including random assignment, some leaders chose not to comply.

We assume there might be many reasons for this. Leaders may have committed to the study in order to receive free program materials.³¹ Leaders may have committed without having the flexibility to accommodate our research needs or without recognizing the importance of maintaining fidelity to the study design. This uncertainty raises an issue about the internal validity of the study⁷ due to the threat to construct validity of the unit.³¹ Although we found the treatment and control conditions to be equivalent on most characteristics at baseline, there is no telling what specific effect these unintended assignment changes would have on outcomes in terms of group/treatment interaction.³¹ However, the result of this probable contamination was an attenuation of differences between the treatment and control means.²⁹ Consequently, our real impact results may, in fact, be larger than our analysis indicates. In future studies, we recommend that all leaders, control as well as treatment, provide implementation data,³¹ and that the method for collection of this data be prospective or by in-person interview with probing questions.

The dose-related results are difficult to explain. The most likely scenario is that our retrospective method of collecting this data was unreliable.

Strengths and other limitations. A major strength of our evaluation study was its randomized, controlled design. Although conducted in the community, we were successful in using the same unit of analysis for randomization, program delivery, and data analysis. Our study was a large-scale trial executed by Cooperative Extension staff who were depending on volunteers without formal research backgrounds to follow a research protocol. For budgetary reasons, the training designed and implemented by the authors for all EFNEP staff involved in the study was not conducted in person, but via a series of conference calls using a written protocol.^{13,23} The conference call method for county staff training may have been inadequate and may have contributed to this lack of fidelity by leaders to our research design. We also did not have the resources to monitor or enforce leader fidelity to assignment of treatment and control conditions. The leaders who volunteered for our study may have reported implementing more of the intervention than they actually did in order to please the researchers. Despite recognizing the value of factor analyzing a large pool of survey items on the evaluation instrument, it was beyond the scope of this study. Another limitation was the forcing of the survey items into the 4 USDA impact indicator "factors" used in the analyses. Future evaluation research should include further testing of

the survey items to improve the instrument's psychometric properties.

Last, we could not afford to hire independent data collection staff. Our county EFNEP staff had a vested interest in having positive outcomes for the study. Although we have no evidence, it is conceivable that county staff lost or falsified documents to enhance outcomes. Although our study had many limitations, we tackled a research quality design not conducted previously by EFNEP. We recommend that similar studies be conducted in other states.

Support for the validity of the USDA impact indicator method for reporting to Congress may be considered the most important outcome of this study. In addition, using a rigorous study design, California Youth EFNEP was shown to be effective in producing favorable gains, although modest, in nutrition knowledge, food selection, and food preparation skills/safety practices. Treatment participants receiving the nutrition education intervention of 7 lessons made greater gains than participants in the control condition. This is the first large-scale evaluation study of Youth EFNEP. These results provide some evidence for continued funding of this nutrition education program for low-income youth.

IMPLICATIONS FOR RESEARCH AND PRACTICE

We have provided an evaluation model that could be duplicated by EFNEP, Food Stamp Nutrition Education (FSNE), and 5-A-Day Power Play in other states and territories interested in documenting program impacts using a research quality design. This research represents the first report to evaluate a state's youth EFNEP intervention using a research-quality protocol. At the same time, it is the first reported estimate of the validity of the USDA evaluation method mandated for EFNEP and other Cooperative Extension programs.

ACKNOWLEDGMENTS

We wish to recognize participating Cooperative Extension County Advisors: Gwendolyn Stanford, Anne Cotter, Yvonne Nicholson, Jeanette Sutherlin, Estella West, Edwina Williams, Eunice Williamson; and EFNEP staff: Elizabeth Gong, Lorri Castro-Aguilera, Margie Alvarez, Suzanne Bernhart, Karen Bayne, Melida Manjarrez, Stella Taylor, Patricia Margolis, Judy D'Innocenti, Trin Vo, Kristin Dufelmier, Martha Weston, Liz Armijo, Pat Garehime. We would like to express our appreciation to the 246 leaders and 5508 students who participated in this study.

REFERENCES

1. Willis W. Cooperative State Research, Education and Extension Service, Expanded Food and Nutrition Education Program Youth Component Impact Data 2001. Available at: <http://www.reeusda.gov/f4hn/efnep/01youth.pdf>. Accessed February 4, 2004.
2. Contento I, Balch GI, Bronner YL, Lytle LA, Maloney SK, Olson CM, Swadener SS. The effectiveness of nutrition education and implications for nutrition education policy, programs, and research: A review of research. *J Nutr Educ*. 1995;27:1-142.
3. Hoelscher DM, Evans A, Parcel GS, Kelder SH. Designing effective nutrition interventions for adolescents. *J Am Diet Assoc*. 2002;102(3):S52-S63.
4. Basen-Engquist K, Parcel GS, Harrist R, Kirby D, Coyle K, Banspach S, Rugg D. The Safer Choices Project: Methodological issues in school-based health promotion intervention research. *J School Health*. 1997;67(9):365-371.
5. Murray DM. Design and analysis of group-randomized trials: a review of recent developments. *Ann Epidemiol*. 1997;7(suppl 7):S69-S77.
6. Levin B. Annotation: Accounting for the effects of both group- and individual-level variables in community-level studies. *Am J Public Health* 1995;85(2):163-164.
7. Campbell DT, Stanley JC. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally; 1966.
8. Hsieh FY. Sample size formulae for intervention studies with the cluster as unit of randomization. *Stat Med*. 1988;8:1195-1201.
9. Townsend MS, Gong E, *Eating Right Is Basic for 9-11 year Olds, Leader's Packet for classroom use, School Program*, Revised and reprinted 1995. 2nd ed. University Publications No. 4148. (Adapted from ERIB, Michigan State Univ.) Available from: University of California, Division of Agriculture and Natural Resources, 6701 San Pablo Ave., Oakland CA 94608-1239; 1-800-994-8849.
10. Contento IR, Manning AD, Shannon B. Research perspective on school-aged nutrition education. *J Nutr Educ*. 1992;24:247-260.
11. Townsend MS, Gong EJ. *Eating Right Is Basic for 9-11 year olds, Advisor's Packet*, University of California, 30 pp training manual, 15 Learning Centers for leaders. 1991. Available from first author.
12. Dennis ML, Boruch RF. Improving the quality of randomized field experiments: Tricks of the trade. In Conrad KJ, ed: *Critically Evaluating the Role of Experiments*. San Francisco: Jossey-Bass, Inc.;1994.
13. Townsend MS, Shilts M. *Report for the Youth EFNEP Evaluation Study, State of California*. Appendix A. 2003. Available from first author.
14. Klare GR. Readability. In: Pearson PD, ed. *Handbook of Reading Research*. New York: Longman, 1984.
15. Klare GR. Assessing readability. *Read Res Quart*. 1974;1:62-102.
16. Litwin MS. *How to Measure Survey Reliability and Validity*. Thousand Oaks, CA: Sage Publications; 1995.
17. Carmines EG, Zeller RA. *Reliability and Validity Assessment*. Newbury Park, CA: Sage Publications; 1979.
18. Willis GB. *Cognitive Interviewing and Questionnaire Design: A Training Manual (Working Paper Series No. 7)*. Hyattsville, MD: Centers for Disease Control and Prevention, National Center for Health Statistics;1994.
19. Townsend MS, Gong E, Murphy S. *Kids Kartoons – Evaluating Adolescent Programs*. Presented at: Nutrition Education for Culturally Diverse Audiences Conference; Feb. 13, 1997; Houston, TX.
20. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. London, UK: Oxford University Press; 1992.
21. Nunnally JC, Bernstein IH. *Psychometric Theory*. 3rd ed. New York: McGraw-Hill, Inc.; 1994:99-100.
22. Pedhazur EJ, Schmelkin LP. *Measurement, Design, and Analysis: An Integrated Approach*. Hillsdale, NJ: Lawrence Erlbaum Assoc.; 1991.
23. Townsend MS, Shilts M. *Youth Evaluation Project FY 1998-1999: Protocol*. September 1998. Available from first author.
24. McGraw SA, Sellers D, Stone E, Resnicow KA, Kuester S, Fridinger F, Wechsler H. Measuring implementation of school programs and policies to promote healthy eating and physical activity among youth. *Prevent Med*. 2000;31:S86-S97.
25. Basch CE, Sliepcevich EM, Gold RS, Duncan DF, Kolbe LJ. Avoiding type III errors in health education program evaluations: a case study. *Health Educ Quart*. 1985;32(4):315-331.

26. Bunch SJ, Gong EJ, Townsend MS, Murphy SP. *A new method for nutrition survey design and automated data entry [abstract]*. Presented at: Food and Nutrition Extension Educators Annual Meeting; July 22, 1997; Montreal, Canada; Abstracts 016.
27. Kerlinger FN. *Foundations of Behavioral Research*. 3rd ed. New York: Harcourt Brace College Publishers; 1992.
28. Flay BR, Phil D. Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Prevent Med*. 1986;15:451-474.
29. Resnicow K, Davis M, Smith M, Lazarus-Yaroch A, Baranowski T, Baranowski J, Doyle C, Wang DT. How best to measure implementation of school health curricula: a comparison of three measures. *Health Educ Res*. 1998;13(2):239-250.
30. Devine JA, Wright JD, Joyner LM. Issues in implementing a randomized experiment in a field setting. In: Conrad KJ, ed. *Critically Evaluating the Role of Experiments*. San Francisco: Jossey-Bass Inc.; 1994.
31. Conrad KJ, Conrad KM. Reassessing validity threats in experiments: Focus on construct validity. In: Conrad KJ, ed. *Critically Evaluating the Role of Experiments*. San Francisco: Jossey-Bass Inc.; 1994.