Evaluation in Practice

A Methodological Approach

SECOND EDITION

Richard D. Bingham

Cleveland State University

Claire L. Felbinger

American University

CHATHAM HOUSE PUBLISHERS

SEVEN BRIDGES PRESS, LLC

NEW YORK · LONDON

PART IV

Reflexive Designs

THE CLASSIFICATIONS OF experimental designs and quasi-experimental designs, the subjects of the previous two parts of this book, are relatively straightforward. That is, they are easy to define. Not so with the subjects of this part, which is concerned with evaluations that use the target group of the program as its own control. Such designs are termed reflexive controls by Peter Rossi and Howard Freeman (1985, 297). Donald Campbell and Julian Stanley considered them to be pre-experimental designs (1963, 171–247). The two types of reflexive designs covered by this part are simple before-and-after studies that are formally known as the one-group pretest-posttest design and the simple time-series design.

For full-coverage programs (i.e., for programs for which most of the population is eligible), it may be impossible to define randomized or constructed control groups or, in fact, to locate nonparticipants. Many programs, for example, are directed at all targets within a specific geographic area. In these cases, the researcher may have no choice but to use reflexive controls.

The essential justification of the use of reflexive controls is that in the circumstances of the experiment it is reasonable to believe that targets remain identical in relevant ways before and after the program. Without the program, pretest and posttest scores would have remained the same.

References

Campbell, Donald T. and Julian C. Stanley. 1963. "Experimental and Quasi-Experimental Designs for Research on Teaching." In *Handbook of Research on Teaching,* edited by N.L. Gage. Chicago: Rand McNally, 171–247.

Rossi, Peter H., and Howard E. Freeman. 1985. *Evaluation: A Systematic Approach.* 3d. ed. Beverly Hills, Calif.: Sage, 297.

CHAPTER 11

One-Group Pretest-Posttest Design

PROBABLY THE MOST commonly used form of reflexive design is the one-group pretest-posttest design, sometimes called comparisons of "before" and "after" data. Not a very powerful design, it is subject to most of the traditional invalidity problems. It is typically used when nothing better can be done. The one-group pretest-posttest design is shown in table 11.1.

The target group is measured before the implementation of the program (O_1) and again after the program is completed (O_2). The difference scores are then examined and any improvement ($O_2 - O_1$) is usually attributed to the impact of the program. The major drawback to this design is that changes in the target may be produced by other events and not the program. The longer the time lapse between the preprogram and postprogram measurements, the more likely it is that other variables besides the program affected the postprogram measurement. Harry Hatry, Richard Winnie, and Donald Fisk outline the conditions under which this design might be applied:

> This design often is the only type that is practical when time and personnel are limited. It is most appropriate (1) when the period covered by the evaluation is short (this making it less likely that non–program related factors will affect the evaluation criteria); (2) when the link between the program intervention and the outcomes being measured is close and direct so no other major events are likely to have had a significant influence on the values measured with the evaluation criteria; or (3) when the conditions measured have been fairly stable over time (and are not, for example, likely to be distorted by seasonal changes), and there is reason to believe such stability will continue. (1981, 28)

Rossi and Freeman give a "hypothetical" example:

> In evaluating the outcome of a nutritional education program testing participants' knowledge of nutrition before and after participation in a three-week set of lectures, the use of reflexive controls is likely

TABLE 11.1
One-Group Pretest-Posttest Design

	Pretest	Program	Posttest
Group E	O_1	X	O_2

> to provide a good measure of the impact of the course because knowledge of nutrition is unlikely to change spontaneously over such a short period of time. (1985, 297–98)

The following article—"Nutrition Behavior Change: Outcomes of an Educational Approach" by Patricia Edwards, Alan Acock, and Robert Johnston—reads much like the Rossi and Freeman example.

READING

Nutrition Behavior Change

Outcomes of an Educational Approach

Patricia K. Edwards
Virginia Polytechnic Institute and State University

Alan C. Acock
Louisiana State University

Robert L. Johnston
Virginia Polytechnic Institute and State University

This study addresses four issues in the evaluation of nutrition education programs: (1) the reliability of knowledge, belief, and behavior scales; (2) the effectiveness of programs targeted to the general public; (3) the longitudinal effects of nutrition education interventions; and (4) the relationship between changes in the cognitive, belief, and behavioral domains. Our findings indicate that reliable knowledge and behavior scales can be developed, but that the internal consistency of belief scales are more problematic. Moreover, improvements in all three domains can be attained with an heterogenous target audience. Although knowledge deteriorates after the course is completed, beliefs remain stable and nutrition behavior continues to improve significantly. Finally, changes in knowledge and beliefs are influential on changes in behavior as a result of the course, but postcourse changes in knowledge and beliefs are not associated with changes in behavior.

AUTHORS' NOTE: *This evaluation was funded by the American Red Cross and the United States Department of Agriculture. We would like to thank the members of the nutrition course development team: Barbara Clarke, Kristen DeMicco, Mary Ann Hankin, Louise Light, Patricia Marsland, Rebecca Mullis, Anne Shaw, and Fred Troutman for their assistance in the development of the outcome measures used in this study.*

A MAJOR RESPONSIBILITY of evaluators is to test the basic theoretical premises underlying the delivery of human service programs (Flay and Best, 1982; Neigher and Schulberg, 1982). With respect to health education programs, numerous studies have examined the assumption that an expansion in pertinent knowledge and positive changes in beliefs is associated with improvements in health-related behavior. Findings from an array of programs—for example, weight reduction (Becker et al., 1977), alcohol use (Goodstadt, 1978), smoking (Thompson, 1978), and breast self-examination (Calnan and Moss, 1984), suggest that the viability of health education in promoting positive health behavior is highly problematic. Moreover, the nature of causal linkages between knowledge, belief, and behavior changes remains in question. Although the consistency model predicts that valid knowledge change is the first stage in a pathway proceeding to changes in beliefs and attitudes—which then culminate in behavior change (Swanson, 1972; Stanfield, 1976; Zeitlan and Formacion, 1981)—some researchers propose alternative causal models. Changes in attitudes may be, in fact, a necessary prior step to the acquisition of health-related knowledge (Mushkin, 1979; Rosander and Sims, 1981), and health-related beliefs can be modified subsequent to changes in behavior to maintain consistency among the affective and behavioral domains (Almond, 1971; McKinlay, 1972).

This body of research is highly relevant to the evaluation of nutrition education programs. However, although a considerable amount of investigation has focused on the magnitude of cognitive, belief, and behavioral outcomes, extant studies suffer from a number of constraints that limit our understanding of the basic premises and potential of nutrition education programs. Firstly, most of the research has examined programs designed for and targeted to specific subpopulations such as low-income mothers (Ramsey and Cloyd, 1979; Rosander and Sims, 1981), primary school children (St. Pierre and Cook, 1981), hospital staff (Looker et al., 1982), and pregnant teenagers (Perkin, 1983). Little is known about the efficacy of programs aimed at a broadly based constituency. Secondly, to our knowledge, there are no studies involving nutrition education programs that report the extent to which positive cognitive, belief, and behavioral changes are sustained over time. Most health education programs are of a relatively short duration, but their objectives—particularly those that are behavior oriented—are not timebound (Hochbaum, 1982; Flay and Best, 1982). Despite long-range expectations, evaluations of the effectiveness of nutrition education programs, and most other health change programs as well, are generally confined to a pretest-posttest design.

A third problem endemic to many nutrition education evaluations deals with the reliability of outcome measures. In some studies the reliability of scales is estimated from tests piloted with samples that are not equivalent to the target audience of the program (i.e., Looker et al., 1982). We cannot assume that because a test is reliable for one population it will be equally reliable for others (Talmage and Rasher, 1981). Other studies, which use internal consistency as a measure of the reliability of their scales, present pretest results only, ignoring the possibility that the scales may have a different level of internal consistency for the posttest stage (i.e., St. Pierre and Cook, 1981; Rosander and Sims, 1981). The difficulty in constructing reliable outcome scales, even when they are administered to fairly homogeneous groups, is demonstrated by the wide variation of reliability coefficients. For example, Sullivan and Schwartz (1981) report a coefficient as low as .00.

Finally most nutrition education program evaluations measure the intended effects of the program in isolation, making assumptions regarding causal linkages between outcome domains, while failing to demonstrate the actual relationships between cognitive, belief, and behavioral changes. Thus, it is impossible to determine if there is, indeed, an empirical association among the knowledge, belief, or behavioral changes that do occur.

This study addresses four issues in the evaluation of nutrition education programs: (1) Can reliable and valid measures of nutrition knowledge, beliefs, and behavior be developed that will enable longitudinal assessment of programs targeted to the general public? (2) Can a program directed to a broad audience be effective in terms of changing nutrition knowledge, beliefs, and behavior? (3) Are positive changes sustained over time? (4) What is the relationship of changes in the cognitive, belief, and behavioral domains? At a time when federal funds for nutrition education programs have been drastically curtailed, these issues are of critical importance in terms of demonstrating the potential of educational interventions as a strategy for promoting positive dietary behavior.

The American Red Cross Nutrition Course Evaluation

This article draws from the results of an evaluation of a nutrition course, "Better Eating for Better Health," developed jointly by the American Red Cross (ARC) and the United States Department of Agriculture (USDA). The primary goals of the course—presently being offered by many of the 3000 ARC chapters nationwide—concern the promotion of nutrition knowledge, positive beliefs, and improved dietary behavior of the general public. The curriculum consists of six two-hour modules that may be presented over a period ranging from two to six weeks. Participant workbooks and supplementary reading materials are provided to each attendee. Prior to conducting a course, ARC instructors are expected to have completed the Nursing and Health Services core curriculum, as well as the Nutrition Instructor Specialty Course. Evaluation was carried out in five separate stages concurrently with the development of the course and included both formative and summative elements. The selected findings reported here are derived from the final stage of the project, a national field test, conducted at 51 ARC chapter sites.[1]

Data Collection Procedures

Course participants were surveyed at three distinct points in time: prior to the beginning of the nutrition course, immediately after the last session of the course, and approximately 10 weeks following completion of the course. Group-administered questionnaires, provided by course instructors who had been trained to implement the surveys, were used to collect baseline (N = 1461) and posttest (N = 1031) data. All individuals attending the first and last sessions of the nutrition course participated in the first two surveys. Although we were unable to use random selection for our sampling procedures, because participation in ARC classes is voluntary, our baseline sample is differentiated along an array of sociodemographic variables.[2] The third participant survey was conducted by means of telephone interviews with a systematically selected subsample of baseline respondents. A total of 248 interviews were attempted to achieve a quota of 200 telephone survey respondents, accounting for a completion rate of 81%. Respondents to the telephone survey are representative of the initial sample along the range of background characteristics.[3]

In order to assess the possible effects of exogenous factors on changes in the nutrition course participant's knowledge, beliefs, and behavior, we solicited volunteers for a nonequivalent control group from individuals who were simultaneously attending other Red Cross courses during the field test. Again, random selection was not possible without severe disruption of ARC chapter activities. The control group completed both the baseline (N = 212) and posttest questionnaires (N = 133). An earlier analysis indicates that although there are some statistically significant sociodemographic differences between the experimental and control groups, these differences are not influential in interpreting the effects of the nutrition education intervention (Edwards et al., 1983a: 15–17, 100–107).

Measurement Techniques

Items for the outcome scales were constructed cooperatively by the course development and evaluation teams, composed of personnel from ARC and USDA, as well as outside consultants. Each item reflects a major knowledge, belief, or behavioral objective of the course.[4] The three outcomes scales were initially tested with a prototype sample of participants during the third stage of the evaluation, which constituted of a "best chance" pilot assessment of course materials and teaching strategies at six ARC chapter sites. (The nutrition course was taught by experienced ARC instructors who were assisted and observed by members of the course development and evaluation teams). Subsequent to this stage of the evaluation, the course development team revised course objectives, materials, and teaching strategies to more adequately coincide with participant needs identified in the evaluation. In addition, the evaluation scales were examined, using factor analysis and alpha reliability to assess the accuracy and validity of each measure. Inadequate items were deleted, and new measures reflecting changes in the nutrition course objectives were added.

The evaluation team conducted a second pilot test of the revised instruments at ten ARC sites during the fourth stage of the evaluation, in a more naturalistic setting, without either observation or supervision. Again, changes were made in the instrumentation to improve the validity and reliability of the items, as well as to ensure that the items represented the fourth-stage modifications in the nutrition course.

The final nutrition knowledge scale consisted of 15 multiple-choice items pertaining to facts about nutrients, sodium, vitamins, food additives, weight loss, and the relationship of disease to nutrition. Respones were coded so that a score of one indicates a correct answer and zero an incorrect response. Nutrition beliefs, operationalized as a dimension of the affective domain (see Fishbein and Raven, 1962), were constructed using a five-category Likert-type scale. The items were coded so that a score of one represents the least positive belief, three denotes the respondent was undecided, and five indicates the most desirable response in terms of the course objectives. The scale included 8 items concerning beliefs that related to the content areas addressed in the knowledge items. The nutrition behavior scale measured the frequency of participant's conduct related to the knowledge and belief questions. The 12 items making up this scale were coded so that one represents the least positive behavior pattern and five the most desirable, according to course objectives. Belief and behavior items were constructed in both positive and negative terms to avoid a response set.

Reliability of Nutrition Outcome Scales

Developing adequate scales to measure nutrition outcomes has been a thorny problem confounding the evaluations of nutrition education programs. Even when a single domain of outcomes is divided into subtests dealing with specific content areas of nutrition knowledge, beliefs, or behaviors, the internal consistency of each subscale is often unacceptable (i.e., Sullivan and Schwartz, 1981). Moreover, as St. Pierre and Cook (1981) illustrate, the reliability of scales may vary considerably among subsamples of a target population.

The Cronbach alpha coefficients for the field test surveys shown in Table 1 demonstrates the problematic nature of achieving reliable nutrition outcome measures. For course participants, the inter-item reliability is only slightly differentiated between the baseline and posttest points on the knowledge and behavior scales. Despite the fact that the scales were administered to a heterogeneous population and include a range of content areas, the reliability coefficients are within an acceptable range. Both scales were examined using principal component factor analysis. All 15 items of the knowledge scale have a positive loading on the first factor, accounting for 21.9% of the variance on the baseline and 25.6% on the posttest. There is a clear first factor on the 11-item behavior scale, explaining 36.0% and 39.1% of the variance on the baseline and posttest, respectively.

Measurement of nutrition beliefs on a single scale had been highly problematic from the start of the project. On our pretest analysis, it was evident that many of the belief objectives of the course were not integrated along a single dimension. Baseline and posttest reliabilities for the eight items that were finally selected are relatively low for course participants. The eight-item scale has a first principal factor explaining 25.8% of the baseline and 29.7% of the posttest variance.

Comparisons between the reliability coefficients for the participant and control group baseline scales show a great deal of consistency for the knowledge and behavior scales. The reliability coefficient for the control group belief scale, however, is considerably lower than that for the treatment group. This difference persists in the comparison of the posttest coefficient. Moreover, the knowledge reliabilities deteriorate for controls on the second test, to a less satisfactory range. The differences between the reliability of scales administered to both groups may, indeed, be a result of the fact that they constitute samples of two separate populations. The participant group, by virtue of electing to take the course, perhaps had a more acute "nutrition awareness," resulting in the higher level of internal consistency of responses. These findings underscore the need to pretest scales with controls, as well as experimentals, when nonequivalent samples are expected.

TABLE 1
Reliability of Nutrition Knowledge, Belief, and Behavior Scales

		Alpha Reliability				
		Course Participants			***Control Group***	
Scale	*Number of Items*	*Baseline*	*Posttest*	*Telephone*	*Baseline*	*Posttest*
Field Test Survey Participants and Controls						
Knowledge	15	.75	.79	—	.72	.62
Beliefs	8	.56	.65	—	.46	.49
Behavior	11	.79	.82	—	.82	.83
Number of respondents		1461	1031	—	212	133
Telephone Survey Participants						
Knowledge	7	.66	.70	.58	—	—
Beliefs	5	.53	.60	.69	—	—
Behavior	7	.78	.77	.71	—	—
Number of respondents		196–199	137–147	196–200	—	—

Note: Number of respondents varies due to missing data. The lower number of respondents included in the posttest results reflects the proportion of participants dropping the class before it was completed, approximately 25%.

In order to keep length of the telephone survey to a minimum, it was necessary to reduce the number of items in each scale for this follow-up survey. Table 1 presents a reexamination of the scales in their reduced form for telephone respondents only. The results show that decreasing the number of items and the pool of respondents has a deleterious effect on scale reliability. Of particular concern here is the deterioration in the interitem coefficients for the knowledge scale used in the telephone survey. Our findings suggest that caution must be taken when developing multiple choice knowledge items for telephone surveys. Although it may be necessary to utilize several different data collection techniques in longitudinal studies, it is evident that each technique must be thoroughly pretested to assure consistent results.

As mentioned previously, prior research has shown that scale reliability can vary among sociodemographic subsets of a sample. When these groups are important in assessing differential outcomes of the program under examination, their responses should be analyzed for scale reliability. Table 2 presents reliabilities for the baseline knowledge, belief, and behavior scales by gender, race, age, education, and income. The findings show that pattern of scale reliability found in the aggregated sample is generally maintained among subsets of respondents. However, none of the scales perform as well for nonwhites and the youngest group of participants. Furthermore, those participants with the lowest level of educational attainment have considerably lower knowledge and belief scale reliabilities than participants who have completed high school or college. Although these findings illustrate the need to test for scale reliability among relevant subsets of a sample, the overall consistency found here does not pose a serious problem in interpreting outcome results for this study.

TABLE 2

Reliability of Baseline Scales by Selected Subsamples of Participants

	Knowledge	*Beliefs*	*Behavior*	*N*
Gender				
Male	.78	.49	.81	177
Female	.74	.56	.78	1269
Race				
White	.72	.57	.80	1105
Nonwhite	.68	.43	.75	298
Age				
<24	.66	.47	.75	219
25–54	.74	.56	.78	855
54+	.76	.60	.78	320
Education				
<HS	.63	.42	.78	143
HS/some college	.69	.51	.77	919
Bachelors +	.69	.56	.82	380
Income				
<$15,000	.67	.50	.77	353
$15,000–29,999	.71	.49	.80	427
$30,000+	.71	.58	.80	468

Note: Variation in the total number for each subsample is due to unreported data.

Effectiveness of Nutrition Education on a Variegated Constituency

The second issue posed in this study concerns the potential of a health education course targeted to a broadly based audience. Table 3 shows that the ARC nutrition course had substantial positive cognitive, belief, and behavioral effects on the overall sample of participants.[5] In contrast, no significant changes are found between the baseline and posttest scores for the control group, despite the fact that their baseline means are not statistically different from those of the treatment group. The key point is that the control group makes no statistically significant improvement and, therefore, the improvement in the treatment group is most reasonably attributed to the course itself.

Our analysis further indicates that positive knowledge, belief, and behavior changes are consistent among subgroups of course participants. There are no statistically significant differences in improvements on the basis of sex, race, and income in the three outcome measures. Although age is a factor that differentiates the degree of positive effects, of the four age groups examined (under 19, 19 to 24, 25 to 54, and 55 or over), all made significant improvements in nutrition knowledge and behavior, and only those participants under 19 years of age did not gain in terms of positive nutrition beliefs. Marital status and education also affect the level of change. Nonetheless, statistically significant and substantial improvements in nutrition knowledge, beliefs, and behavior are found within each category of these two variables. Our results show that although sociodemographic variables have selected influences on how much participants gain from the course, they do not differentiate to the extent that certain groups fail to improve their nutrition knowledge, belief, or behavior after taking the course.[6]

One reservation must be noted here, however. Participation in the ARC nutrition course is, by design, voluntary. There is no reason to expect that we could find the same dramatic positive results with a group of participants who had not been motivated to participate in the course. Our reservations may, however, be tempered by the fact that the reasons for participating in a health-related Red Cross class are extremely complex. Analysis of open-ended responses illicited from course participants indicates that though many of the respondents decided to attend the course to improve their own dietary behavior, others were motivated to do so because the course was job-related or provided an opportunity for social interaction.

Are Positive Changes Sustained Over Time?

Our third concern in evaluating a nutrition education program relates to the longitudinal effects of the intervention. Table 4 presents the mean changes in nutrition knowledge, beliefs, and behavior for the modified instruments used in the telephone survey undertaken approximately ten weeks after the course was completed. Our analysis only in-

TABLE 3

Changes in Nutrition Knowledge, Beliefs, and Behavior: Immediate Effects

	Participants				***Controls***			
Scale	*Means (N = 883)*	*S.D.*	*Change*	*Probability (2-tailed)*	*Means (N = 104)*	*S.D.*	*Change*	*Probability (2-tailed)*
Knowledge								
Baseline	8.91	3.24	1.96	.000	8.09	3.36	.10	NS
Posttest	10.87	3.24			8.19	3.14		
Beliefs								
Baseline	27.49	4.17	3.29	.000	26.82	3.67	–.10	NS
Posttest	30.78	4.38			26.72	3.70		
Behavior								
Baseline	36.00	7.76	5.05	.000	36.08	8.01	.94	NS
Posttest	41.05	7.10			37.02	7.91		

Note: Possible ranges for each scale are: Nutrition knowledge, 0–15; Nutrition beliefs, 5–40; Nutrition behavior, 5–55. It should be noted that the absolute degree of change cannot be compared across the scales due to differences in the length and coding of the instruments. The probabilities are based on individual T-tests.

TABLE 4
Telephone Survey Comparison: Follow-Up Effects

	Knowledge				*Beliefs*				*Behavior*			
	N	*Means*	*S.D.*	*Change*	*N*	*Means*	*S.D.*	*Change*	*N*	*Means*	*S.D.*	*Change*
Baseline to posttest	147	5.11 5.93	1.80 1.50	.82*	136	16.62 19.15	3.05 2.98	2.53*	132	22.32 25.48	5.50 4.82	3.16*
Posttest to telephone	147	5.93 5.69	1.50 1.39	-.24*	137	19.12 18.71	2.99 3.18	-0.41	132	25.56 26.63	4.79 4.39	1.07*
Baseline to telephone	200	5.21 5.69	1.75 1.42	.48*	197	16.67 18.62	3.06 3.20	1.95*	193	21.84 26.33	5.49 4.84	4.49*

Note: Possible ranges for each scale are: Nutrition knowledge, 0–7; Nutrition beliefs, 5–25; Nutrition behavior, 5–35. Because some of the original items were deleted in the telephone survey instruments, these scores are not comparable to the change scores in Table 2.
*$p < .05$ based on individual T-tests.

cludes participants who had been involved in the follow-up telephone survey. Because we wanted to acquire further information from individuals who had, for some reason, dropped the course, the attrition rate is reflected in the lower number of cases available for analyses involving posttest responses.

The first two rows in Table 4 provide the change when baseline and posttest scores are compared. These changes are substantial and significant for all three measures in a positive direction. Thus, the outcome results for this subsample of participants are consistent with the full complement of participants. We expected a significant drop-off in the scores between the posttest and the telephone survey. In the weeks that elapsed between the two measurements, a great deal of information and the motivation to change behavior could be lost. Indeed, there is a significant reduction in the score on the nutrition knowledge scale. We can conclude that although participants learn a great deal initially, a substantial proportion of this new knowledge is lost soon after completing the course. In contrast, although there is some loss in terms of positive beliefs, the deterioration is not statistically significant. What is, perhaps, most surprising is the significant improvement in the quality of behavior at the end of the ten week period. Finally, looking at the comparison of the baseline and telephone responses in the last two rows of Table 4, we can see that, despite the drop in knowledge subsequent to completion of the course, knowledge, belief, and behavior improvements persist.[7]

We might speculate that the initial improvement in nutrition knowledge provides a cognitive influence in the beliefs that participants have about good nutrition. Even after the students lose some of their specific knowledge, the improved beliefs are retained. Perhaps behavioral changes become self-reinforcing elements in the participant's lifestyle.

The Relationship of Cognitive, Belief, and Behavioral Domains

Table 5 presents the correlations between changes in the outcome domains for two data sets. The first column of correlations shows the changes between the baseline and the posttest. The second column does the same for the changes between the posttest and subsequent telephone interview. These correlations refer to change scores rather than the actual scores on the scale themselves. For example, the correlation of .21 between knowledge and belief means that the more a person improves their knowledge between the baseline and the posttest, the more they also improved their beliefs.

The changes between the baseline and the posttest are positively correlated across all three domains. These correlations are modest, but all are statistically significant. The more a person improves his or her knowledge, the more beliefs ($r = .21$; $p \leq .5$) and behavior ($r = .15$; $p \leq .5$) also improve. Similarly, the more one's beliefs improve, the more one's behavior improves ($r = .26$; $p \leq .5$).

Analyzing the pattern of these correlations, we can see that changes in beliefs may have a greater influence on improvements in behavior ($r = .26$) than do changes in knowledge alone ($r = .15$). Although both of these correlations are statistically significant by themselves, the difference between them does not achieve statistical significance ($p = .11$, one-tail). Therefore, we hesitate to generalize these results beyond this particular data set. Nonetheless, these results do suggest that educational interventions should address changing beliefs, along with providing relevant knowledge. At the very least, these results emphasize that changing beliefs are an important covariate of changing behavior.

Examining the correlations that appear in the last column of Table 5 provides very different results. These correlations reflect the relationship between changes in knowledge, beliefs, and behavior that occur from the posttest to the subsequent telephone interviews. Earlier we demonstrated that behaviors continue to improve after the completion of the course, whereas knowledge deteriorates significantly and beliefs deteriorate somewhat, but not significantly. The fact that none of these correlations is statistically significant means that changes after the course in one domain are independent of changes in the two other domains. This is especially true for nutrition behavior. Nutrition behavior improves after the course, regardless of what happens to the participant's knowledge and beliefs following the course. Thus, a subject who forgets much of what is learned from the course is nearly as likely to continue improving nutrition behavior as a person who remembers all of the nutrition information and belief material.

TABLE 5

Correlation of Changes in Nutrition Knowledge, Belief, and Behavior Scales

Change	*Baseline to posttest (N = 883)*	*Posttest to telephone (N = 134)*
Knowledge/belief	0.21*	–0.08
Knowledge/behavior	0.15*	–0.10
Belief/behavior	0.26*	–0.05

*p < .05.

These results indicate that change in knowledge and especially change in beliefs are important to produce the initial changes in behavior (baseline to posttest). Just as important, however, the maintenance and enhancement of improved nutrition behavior is largely independent of how much of the knowledge and belief information is retained subsequent to the course.

Conclusions

At the beginning of this article we raised four questions about the measurement of nutrition outcome variables; what can be accomplished on a large and heterogeneous population; the ability to sustain changes over time; and the interdependence of changes in cognitive, belief, and behavioral domains.

This study has the advantage of being a formative evaluation. This allowed two systematic pretests of the measurement scales prior to the actual implementation of the evaluation. There is an additional advantage in terms of having the opportunity to share the empirical results of the pretest with the program personnel and combine the empirical and expert opinions in developing the final scales. Unfortunately, much evaluation research will not have these advantages.

It was possible to develop a highly reliable scale to measure nutrition behavior. This high reliability is consistent using the group-administered long form of the questionnaire (11 items) for the case of both the baseline and the posttest in both the treatment group and in the control group. The long-form result is also consistent with the short form of the scale (7 items) used in the telephone interviews. This consistency is evident, even though fewer items were used and the method of measurement was changed from group-administered to telephone interview.

Beliefs are consistently problematic. The reliability for the control group is poor. Interestingly, for the experimental group there is improvement in the reliability between the baseline and the posttest and even more improvement moving from the posttest to the telephone survey. Perhaps there is a tendency to organize one's beliefs about nutrition as a result of the course. Thus, the respondent's nutrition beliefs not only improve but become more integrated due to the systematic exposure to nutrition information in the course.

Knowledge is reliable, but the reliability appears to drop off on the telephone follow-up. Perhaps this is because knowledge is the most likely of the three domains to deteriorate after the course is completed. Thus, nutrition knowledge not only deteriorates, but the integration of that knowledge in a coherent fashion also drops off.

A general conclusion regarding the reliability of the scales is that it is possible to develop reasonably reliable sales even though we are applying them to a heterogeneous population and using different methods to administer them. Impressively, the telephone interviews do nearly as well as the group administered questionnaire method even though the telephone interviews have fewer items in the scales.

The second major question concerns the ability to produce changes in a broad audience. Our results show that progress is substantial with a diverse population. However, as we have dealt with a volunteer population, we do not know if such impressive improvements in knowledge, belief, and behavior are possible with a less motivated, nonvoluntary population.

The third issue concerns the ability to sustain changes over time. What happens after the course was completed? As our results show, there is a significant deterioration in nutrition knowledge, a slight (nonsignificant) decrease in the quality of beliefs, and an actual improvement in nutrition behavior after the course is completed.

Although we had hoped behavior would be sustained after the course, we were surprised that it actually improved significantly. One possible explanation for this improvement can be gained by examining some of the classic studies conducted by social psychologists on the effect of fear on changes in attitudes and behavior. Leventhal (1980) reports great inconsistency in the results of the use of fear in change programs. He suggests that fear produces two possible effects. One of these is a realistic fear that can be controlled by directed changes in behavior. Thus, when people are scared and are told what specific behavior will mitigate this fear, they are likely to change their behavior dramatically.

The nutrition course did not intend to produce high levels of fear among the participants. Still, the substantial changes in knowledge and beliefs could be expected to make people fearful of practicing poor nutrition behavior. On the other hand, proper nutrition conduct could be expected to give a very positive feeling because it mitigates concerns developed because of the improved knowledge and beliefs.

This argument means that proper dietary behavior may become a self-reinforcing behavior. For example, proper diet makes people feel good because they do not need to be concerned about their health (at least as far as nutrition is a health issue). Even after the participant forgets some of the details they learned in the course, the good feeling associated with proper nutrition behavior retains its self-reinforcing property and continues to improve.

If this argument is valid, then the findings have far-reaching implications for change programs. The creation of moderate levels of fear based on improved knowledge and beliefs combined with clear guidelines on specific behavior to control such fear can make the behavior self-reinforcing. Once the behavior becomes self-reinforcing, it can continue to improve long after the completion of the intervention and in spite of a deterioration in the level of specific knowledge and beliefs.

The final major issue of this article is the relationship among the three domains (cognitive, belief, and behavioral) of nutrition. Some interventions focus on only one or two of these domains, however, we have demonstrated that all three are interrelated. In particular, positive beliefs appear to be even more important improvement than knowledge as a prescursor of changes in behavior. Although changes in knowledge and beliefs are influential on changes in behavior as a result of the course, postcourse changes (drop-off) in beliefs and knowledge are not associated with changes in behavior. This is further evidence that nutrition behavior can become self-reinforcing after the completion of the course.

Notes

1. This article reports only selected findings from the field test, which included a more comprehensive analysis of the relationship between participant attributes, program delivery variables, (e.g., instructor training and experience) exogenous factors, and a variety of program outcomes, such as attendance, extent of involvement, level of difficulty of the course, and participant satisfaction with course content, materials, and teaching strategies. A description of the evaluation model and results of the analysis have been reported elsewhere (Edwards, 1984; Edwards et al., 1983a; Edwards et al., 1983b). Data from two additional field test surveys directed to course instructors and chapter administrators are not included in this analysis.
2. For example, gender: female (88%); age: <25 years (16%); 25–34 (24%); 35–54 (39%); 55+ (21%); marital status: married (54%), single (25%), divorced/widowed (21%); ethnicity: white (77%), black (17%), Hispanic (4%), other (2%); household income: <$15,000 (28%); $15,000–34,999 (44%); $35+ (28%); employment status; full time (43%), part time (15%), not employed (42%); education <high school (10%), high school graduate (28%), some college (36%), college graduate (26%).
3. There are no statistically significant differences among the baseline sample and telephone respondents in terms of sex, age, educational attainment, income, race, marital status, mean household size, or participant's perception of their personal health status.
4. To illustrate, an objective of session 3 was to clarify knowledge of health problems related to excess sugar consumption and suggest strategies for reducing sugar in the diet. Examples of survey questions used to measure this objective within each outcome domain are shown below.

   Knowledge: The most common disease in the United States related to excessive sugar consumption is:

   (1) diabetes

   (2) tooth decay

   (3) obesity

(4) cirrhosis

(5) gastrointestinal disorders

Beliefs: Honey and molasses are better for you than table sugar (strongly disagree, disagree, undirected, agree, strongly agree).

Behavior: (I) eat fruits canned in their own juice without sugar added rather than fruits canned in syrup (never, rarely, sometimes, usually, always).

The complete sets of scale items can be obtained from the authors.

5. Analysis of the effects of the course showed that sociodemographic factors explained only a small proportion of the variance on those outcome measures (Edwards et al., 1983a).
6. Multivariate analyses of the effect of process and participant perception variables (e.g., extent to which the course design was followed, team versus individual instruction and participant assessments of course content, activities, length, difficulty, and quality of instruction) on the three outcome domains are reported elsewhere (Edwards, 1984; Edwards et al., 1983a).
7. Because 25% of the telephone sample had dropped the course, these results may underestimate the longer range positive effects of the intervention for participants who attended most of the course.

References

Almond, R. (1971) "The therapeutic community." *Scientific Amer.* 224: 34–42.

Becker, M. H., L. A. Maiman, J. P. Kirscht, E. P. Haefner, and R. H. Drachman (1977) "Dietary compliance: a field experiment." *J. of Health and Social Behavior* 18: 348–366.

Calnan, M. W. and S. Moss (1984) "The health belief model and compliance with education given at a class in breast self-examination." *J. of Health and Social Behavior* 25: 198–210.

Edwards, P. K. (1984) "The American Red Cross nutrition course: findings from the field test," pp. 575–586 in *Agricultural Outlook 85*. Washington, DC: U.S. Department of Agriculture.

Edwards, P. K., A. C. Acock, R. L. Johnston, and K. L. Demicco (1983a) *American Red Cross Nutrition Course Field Test: Technical Report.* Washington, DC: American Red Cross.

Edwards, P. K., R. Mullis, and B. Clarke (1983b) "A comprehensive model for evaluating innovative nutrition education programs." Paper presented at the Evaluation Research Society meetings, Baltimore, MD.

Fishbein, M. and B. H. Raven (1962) "The AB scales: an operational definition of beliefs and attitudes." *Human Relations* 15: 35–44.

Flay, B. R. and J. A. Best (1982) "Overcoming design problems in evaluating health behavior programs." *Evaluation and the Health Professions* 5: 43–69.

Goodstadt, M. (1978) "Alcohol and drug education: models and outcomes." *Health Education Monographs* 6: 263–279.

Hochbaum, G. W. (1982) "Certain problems in evaluating health education." *Amer. J. of Public Health* 6: 14–20.

Leventhal, H. (1980) "Findings and theory in the study of fear communications," pp. 120–186 in L. Berkowitz (ed.), *Advances in Experimental Social Psychology,* Vol. 5. New York: Academic.

Looker, A., S. Walker, L. Hamilton, and B. Shannon (1982) "Evaluation of two nutrition education modules for hospital staff members." *J. of the Amer. Dietetic Assn.* 81: 158–163.

McKinlay, J. B. (1972) "Some approaches and problems in the study of the use of services: an overview." *J. of Health and Social Behavior* 14: 115–151.

Mushkin, S. J. (1979) "Educational outcomes and nutrition," pp. 269–302 in R. E. Klein et al., (eds.) *Evaluating the Impact of Nutrition and Health Programs.* New York: Plenum.

Neigher, W. D. and H. C. Schulberg (1982) "Evaluating the outcomes of human service programs: a reassessment." *Evaluation Rev.* 6: 731–752.

Perkin, J. (1983) "Evaluating a nutrition education program for pregnant teenagers: cognitive vs. behavioral outcomes." *J. of School Health* 53: 420–422.

Ramsey, C. E. and M. Cloyd (1979) "Multiple objectives and the success of educational programs." *J. of Nutrition Education* 11: 141–145.

Rosander, K. and L. S. Sims (1981) "Measuring effects of an affective-based nutrition education intervention." *J. of Nutrition Education* 13: 102–105.

St. Pierre, R. G. and T. D. Cook (1981) "An evaluation of the nutrition education and training program." *Evaluation and Program Planning* 4: 335–344.

Stanfield, J. P. (1976) "Nutrition education in the context of early childhood malnutrition in low resource communities." *Proceedings of the Nutrition Society* 35: 131–144.

Sullivan, A. D. and N. E. Schwartz (1981) "Assessment of attitudes and knowledge about diet and heart disease." *J. of Nutrition Education* 13: 106–108.

Swanson, J. E. (1972) "Second thoughts on knowledge and attitude effects upon behavior." *J. of School Health* 42: 362–363.

Talmage, H. and S. P. Rasher (1981) "Validity and reliability issues in measurement instrumentation." *J. of Nutrition Education* 13: 83–85.

Thompson, E. L. (1978) "Smoking education programs 1960–1976." *Amer. J. of Public Health* 68: 250–257.

Zeitlan, M. F. and C. S. Formacion (1981) *Nutrition Education in Developing Countries: Study II, Nutrition Education.* Cambridge, MA: Oelgeschlager, Gunn, & Hain.

Patricia K. Edwards is Associate Professor of Urban Affairs in the College of Architecture and Urban Studies at Virginia Polytechnic Institute and State University. She has had experience evaluating international, national, and local programs dealing with public health, housing, recreation, and community development.

Alan C. Acock is Professor of Sociology, Rural Sociology, and Graduate Director in the Sociology Department at Louisiana State University, Baton Rouge. His current research concerns using LISREL to isolate a family effect from individual effects, reviewing intergenerational relations research, and using personal computers in content analysis.

Robert L. Johnston is completing his doctorate in Sociology at Virginia Polytechnic Institute and State University and is an Adjunct Instructor at Roanoke College. His recent research projects include the analysis of labor process development in the post-World War II era.

Explanation and Critique

The first thing a student is likely to say on reading the Edwards, Acock, and Johnston article is this: "This is not an example of the one-group pretest-posttest design; it had a comparison group. It is really a quasi-experimental design." But is it? Look at the article another way.

Assume that you are taking a course in program evaluation at American University and Professor Felbinger is your instructor. Down the hall from you is a class in astrophysics. Professor Felbinger decides to find out how much you learn about evaluation research in the course, so she gives you a comprehensive 200-question multiple choice test on evaluation research on the first day of class. The class performs miserably. She gives you the same test on the last day of class and you all do quite well. She also gives the same test on the first and last days of class to the physics students. They perform miserably on both the pretest and the posttest. Has the use of a comparison group—the physics students—somehow magically transformed Professor Felbinger's evaluation from a reflexive design to a quasi-experimental design? Of course it has not. There is no reason in the world to expect that the population in general, including physics students, would show a statistically significant increase in their knowledge of program evaluation practices during the course of a semester.

Remember, when identifying a comparison group, to carefully match characteristics of the comparison group with those in the experimental group. If, for example, your experimental group consists of cities, make the matches on the basis of similar size, revenue base, services delivered, and so forth. Select groups of people on normal demographic variables so that the groups are similar in all possible *relevant* respects except the treatment. In the Edwards, Acock, and Johnston article, there is no reason to expect that a comparison group of individuals attending other Red Cross courses (e.g., life saving) would either be similar in relevant respects to those interested in nutrition training or would show a statistically significant improvement in their knowledge of nutrition during the time frame that the nutrition course was being offered. One might call this the use of a pseudo-control group—at most, it controls for the effects of history. It certainly is not a systematically chosen comparison group.

In reality the Edwards, Acock, and Johnston article presents an example of the one-group pretest-posttest design. Volunteers for a "Better Eating for Better Health" course consisting of six two-hour modules (presented over a period ranging from two to six weeks) were tested at three points: before the beginning of the course, immediately after the last session of the course, and approximately ten weeks later. The results showed a statistically significant change in knowledge, beliefs, and behavior (table 3 of the article). Pretty simple, is it not? There is no reason to expect that exogenous factors caused these changes, nor is there any reason to cloud up a simple (and appropriate) design with a pretense of elegance.

The article can be criticized from the standpoint of the comparison group, but the evaluation has its strengths. In particular, the authors handled the problem of measuring the outcome of their nutritional education program very well. They were careful in the selection of their measurement items and took the time to compute reliability coefficients. They also examined reliabilities for knowledge, belief, and behavior scales by gender, race, age, education, and income.

Their analysis indicated significant changes in participants' knowledge, belief, and behavior as a result of the course. Furthermore, there were no significant improvements based on gender, race, or income. Age, marital status, and education, however, did affect the level of change.

Edwards, Acock, and Johnson were quite thorough in this evaluation in that they were also interested in how well the results held up over time. To test this, they used a telephone survey of participants ten weeks after the course was completed—expecting a significant drop in the scores between the original program posttest and the telephone survey. They found a significant reduction in the score on the nutrition knowledge scale, a smaller reduction in terms of positive beliefs (the deterioration was not statistically significant), and a significant improvement in the quality of nutrition behavior at the end of the ten weeks. Thus, despite a drop in knowledge, changes in belief and behavior brought about by the course persist. The authors conclude:

> The creation of moderate levels of fear based on improved knowledge and beliefs *combined* [emphasis in original] with clear guidelines on specific behavior to control such fear can make the behavior self-reinforcing. Once the behavior becomes self-reinforcing, it can continue to improve long after the completion of the intervention and in spite of a deterioration in the level of specific knowledge and beliefs.

Without the second posttest, such conclusions would never have been possible.

References

Campbell, Donald T., and Julian C. Stanley. 1963. "Experimental and Quasi-Experimental Designs for Research on Teaching." In *Handbook of Research on Teaching,* edited by N.L. Gage. Chicago: Rand McNally, 171–247.

Rossi, Peter H., and Howard E. Freeman. 1985. *Evaluation: A Systematic Approach.* 3d ed. Beverly Hills, Calif.: Sage.