DOCUMENT RESUME

ED 320 856 JP 032 099

TITLE	Smoking Cessation. Program Evaluation Handbook.
INSTITUTION	IOX Assessment Associates, Culver City, CA.
SPONS AGENCY	Centers for Disease Control (DHHS/PHS), Atlanta, GA.; Public Health Service (DHHS), Rockville, MD. Office of Disease Prevention and Health Promotion.
PUB DATE	88
NOTE	130p.; For the other titles in this series, see SP 032 101-105 and SP 032 107.
AVAILABLE FROM	IOX Assessment Associates, 5420 McConnell Ave., Los Angeles, CA 90066 ($17.95).
PUB TYPE	Guides - Non-Classroom Use (055)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	*Attitude Measures; *Evaluation Criteria; Evaluation Methods; *Program Evaluation; *Psychometrics; Questionnaires; *Smoking
IDENTIFIERS	*Smoking Cessation

ABSTRACT

Intended as a resource for individuals wishing to evaluate smoking cessation programs, this handbook, one in a series of seven, provides a collection of measuring devices that can improve the quality of such evaluations. Chapter 1 introduces the handbook's contents and outlines evaluation related issues specific to smoking cessation programs. Chapter 2 introduces the key operations involved in program evaluation, emphasizing the role of assessment instruments in the gathering of information needed for defensible evaluations. Chapter 3 contains the measures designed to be used in the evaluation and design of smoking cessation programs. These measures deal with behavior, knowledge, skills, and affective outcomes. Each measure is introduced by a brief statement on the purpose of the assessment instrument, as well as the procedures for administering, scoring, and analyzing the resulting data. Chapter 4 describes how technical appraisals of the handbook's measures can be carried out. The three appendices contain lists of amplified content descriptors for use in updating the information in the guide, an explanation of informed consent procedures, and an annotated bibliography. (JD)

PROGRAM EVALUATION HANDBOOK

Smoking Cessation

A collection of practical guidelines and measures for smoking cessation programs

PROGRAM EVALUATION HANDBOOK

SMOKING CESSATION

Prepared for

The Center for Health Promotion and Education
United States Centers for Disease Control

The Office of Disease Prevention and Health Promotion
Office of the Assistant Secretary for Health
United States Department of Health and Human Services

by

IOX Assessment Associates
P.O. Box 24095
Los Angeles, CA 90024-0095

1988

TABLE OF CONTENTS

Preface

In recent years, health educators have increasingly recognized that systematic evaluation can help them appraise and improve their programs. For this potential to be realized, however, effective mechanisms for gathering relevant data are required. In the past, critical information about a program's effects was not collected in some instances because suitable measures for gauging those effects were lacking. The purpose of this handbook is to rectify, at least in part, this deficiency in the evaluation of health education programs dealing with smoking.

This book is one of seven health education evaluation handbooks resulting from a project jointly initiated in 1980 by the United States Centers for Disease Control (CDC) and the Office of Disease Prevention and Health Promotion (ODPHP) of the Office of the Assistant Secretary for Health. The handbook is not intended to be prescriptive or all-inclusive. Those who evaluate smoking cessation programs should regard the handbook as only a resource, that is, a collection of assessment tools that may be of use in program evaluation. The extent to which the handbook will actually be useful depends chiefly on the extent to which it contains assessment tools that correspond to the evaluation needs of a particular smoking cessation program.

Handbook Development

This handbook has been created by IOX Assessment Associates (IOX), selected competitively on the basis of responses to a governmentally issued request for proposals. IOX was to collect and develop program evaluation measures for critical behavioral, knowledge, and affective outcomes in the area of smoking cessation. Three panels of experts played prominent roles in the creation of this handbook. A **Handbook-Development Panel**, consisting of six experts familiar with smoking cessation programs or their evaluation, guided the initial development of the handbook. The Handbook-Development Panel identified important outcomes for smoking cessation programs. IOX staff, drawing on the advice of panelists, then developed assessment instruments to assess panel-identified program outcomes. The names and affiliations of the Smoking Cessation Handbook-Development Panelists are provided below:

Handbook-Development Panel

Dr. Peter A. Cortese
California State University
Long Beach, California

Dr. Brian G. Danaher
Brian G. Danaher & Associates, Inc.
Pasadena, California

Dr. Nancy Doyle
American Lung Association
New York, New York

Dr. Gilbert Sax
University of Washington
Seattle, Washington

Dr. Betty Tevis
American Heart Association
Dallas, Texas

Dr. Jane Zapka
University of Massachusetts
Amherst, Massachusetts

The Handbook-Development Panel met at the beginning of the project in order to isolate the chief outcomes that smoking cessation programs could reasonably be expected to promote. Preliminary statements reflecting these outcomes were identified by the panelists. These preliminary outcome statements were refined by IOX staff and mailed to the panelists and other interested specialists, all of whom rated the importance of each statement. The list of high-priority outcomes that resulted was used to guide the selection and development of the original handbook's measures.

All newly developed measures were mailed to the panelists for review. In addition, all of these measures were tried out with small groups of respondents. The measures were revised based on the informal tryouts and the panelists' review comments. All of the new measures were also reviewed by IOX staff in an effort to eliminate any potential ethnic, gender, religious, or socioeconomic bias.

A completed version of the smoking cessation handbook was delivered to the government in 1983. Several thousand copies of the handbook were released by CDC and ODPHP to health educators throughout the nation.

Handbook Revision

Subsequent to the initial distribution of the handbook, CDC issued, in concert with ODPHP, a second request for proposals which led to the comprehensive revision of the existing smoking cessation handbook. To guide the review and revision of the smoking cessation handbook, a **Handbook-Revision Panel** was constituted. Members of the panel were selected because of their dual expertise in (a) the field of smoking cessation and (b) measurement of the outcomes sought by smoking cessation programs. Members of the Handbook-Revision Panel and their affiliations are listed below:

Handbook-Revision Panel

Dr. J. Alan Best
University of Waterloo
Ontario, Canada

Dr. C. Anderson Johnson
University of Southern California
Los Angeles, California

Dr. Edward Lichtenstein
University of Oregon
Eugene, Oregon

Dr. Ian Newman
University of Nebraska
Lincoln, Nebraska

Dr. Jonathan Fielding
University of California
Los Angeles, California

Dr. Patricia Mullen
University of Texas
Houston, Texas

Dr. Donald Iverson
University of Colorado
Denver, Colorado

Dr. Thomas Glynn
National Cancer Institute
Bethesda, Maryland

The Handbook-Revision Panel met on two occasions. In these meetings, panelists reviewed the contents of the initial version of the smoking cessation handbook, particularly

its measures, and suggested deletions, modifications, or additions. Panelists also provided guidance regarding ways of making the handbook more usable to practitioners. During both of these meetings, the panelists were attentive to the accuracy of the handbook's contents. Considerable content, in the measures as well as the introductory materials, was revised or deleted on the basis of panelists' suggestions.

Overall Guidance

A third panel, the **Project Advisory Panel**, provided overall guidance to IOX staff during the final three years of the project. These individuals offered technical counsel and strategic advice during the revision of all handbooks. Members and affiliations of the Project Advisory Panel are listed below:

Project Advisory Panel

Dr. Peter A. Cortese
California State University
Long Beach, California

Dr. William L. Haskell
Stanford University
Stanford, California

Dr. Lawrence W. Green
Henry J. Kaiser Family Foundation
Menlo Park, California

Dr. Jonathan E. Fielding
U.S. Corporate Health Management
and University of California
Los Angeles, California

Acknowledgments

The project that led to the creation of this handbook was funded by the CDC and ODPHP. Dr. Walter J. Gunn of CDC conceptualized the project and supplied technical guidance throughout its first phase. During this time, Dr. Diane Orenstein of CDC as well as Dr. Donald Iverson and Dr. Patricia Mullen, both of ODPHP, served as project officers.

During the project's second phase, that is, the revision of the handbook, Dr. Orenstein of CDC continued to serve as project officer along with Dr. Joel Kavet, Dr. Gregory Christenson and Mr. James Harrell of ODPHP.

As the handbook progressed, numerous health educators throughout the nation offered their insights regarding the handbook's contents. Without their expert assistance, development of this volume would have been impossible.

IOX Assessment Associates
July, 1988

CHAPTER ONE

A Resource for the Evaluation of Smoking Cessation Programs

A Resource for the Evaluation of Smoking Cessation Programs

This handbook is intended to help those individuals who wish to evaluate health education programs dealing with smoking cessation. More specifically, the handbook provides a series of measuring devices that, if selected and used judiciously, can improve the quality of such evaluations. As a consequence, not only will the technical quality of the program evaluation be improved, but any program-related decisions based on the evaluation's results are apt to be more defensible.

An Evidence-Oriented Era

In recent years, educators have experienced substantially increased pressures to produce evidence that their programs are functioning effectively. In contrast to an earlier era when it was widely thought that most educational programs were worth the money they cost, today's educators find that they are constantly called on to justify the effectiveness of their programs.

The kinds of evidence that health educators have been required to assemble regarding program effectiveness have, almost without exception, involved the use of various kinds of assessment instruments. Consonant with that requirement, this handbook contains numerous tests and inventories designed to secure the evidence needed to judge the effectiveness of smoking cessation programs. The handbook's measuring instruments were created specifically to assess important goals of the most common types of smoking cessation programs offered for adults (in industrial or clinical settings) and for children (in school-related programs).

The handbook, accordingly, makes available to those who operate smoking cessation programs the assessment tools by which the effectiveness of such programs can be determined. The evidence of program effectiveness currently being demanded of smoking cessation personnel can, therefore, be provided by appropriate use of the handbook's assessment instruments. Moreover, as will be indicated shortly, appropriate use of the handbook's numerous assessment devices can substantially improve the *design* of smoking cessation programs.

Measurement and Program Design

Historically, assessment devices have been thought of as instruments to be used *after* a program was concluded. Teachers, for example, have traditionally administered tests *after* instruction was over in order to grade students. However, even though assessment instruments have often been post-instruction creations of instructors, such instruments can make important—often overlooked—contributions to the original design of an instructional program. Properly developed assessment tools, in fact, can contribute to program design in two significant ways.

First, because assessment instruments are typically intended to measure outcomes of interest, such assessment instruments provide program personnel with a range of potential

outcomes. An increased range of possible program outcomes generally leads to the *selection of more defensible outcomes* for health education programs. To illustrate, there may be an assessment instrument dealing with an attitudinal dimension that, were it not for the measuring instrument's availability, might have been overlooked by the program staff. Stimulated by the assessment tool's availability, however, the program staff can add the attitudinal dimension to the program's targeted outcomes.

A second program-design dividend of properly constructed assessment tools is that they *clarify intended program outcomes* and, thereby, make possible the provision of more on-target program activities than would have been the case had such clarification not been present. To illustrate, suppose that program personnel intend to feature in their evaluation an assessment device focused on the knowledge of the effects of smoking on society. By becoming familiar with the composition of that assessment tool, the program staff can be sure to incorporate critical facts about those effects in their instructional program. Provision of appropriate instructional practice for participants need not reflect "teaching to the test" in the negative sense that instructors coach students for specific test items. Instead, providing relevant knowledge so that program participants attain the program's intended outcomes constitutes an efficient and effective, research-supported form of instruction.

To review, then, the measuring instruments provided in this handbook are intended to assist those who design and those who evaluate smoking cessation programs. With respect to program evaluation, the measures will yield evidence by which to improve programs as well as determine program effectiveness. With respect to program design, the measures provide a menu of potential program options and, once having been selected, enhanced clarity regarding the nature of the outcome(s) sought.

What the Handbook Contains

There are several key ingredients in this handbook. It should, therefore, prove helpful to readers if the handbook's major sections are presented. Briefly, then, here is a description of the handbook's major components:

Introductory information. In Chapter One, an introduction to the handbook is provided. Because the handbook is intended to be used with smoking cessation programs, the chapter concludes with a brief discussion of evaluation-related issues specific to health education programs dealing with smoking cessation.

Program evaluation essentials. Although a number of people who use this handbook will already be familiar with the nature of program evaluation, many handbook users will not be well versed in the conduct of program evaluations. Accordingly, in Chapter Two, an introduction is provided to the key operations involved in program evaluation. Although space limitations preclude a detailed exposition of all aspects of program evaluation, emphasis is given to the role that assessment instruments play in the gathering of information needed for defensible evaluations.

Assessment instruments. Chapter Three contains the handbook's most important components, namely, the measuring tools designed to be used in the evaluation and design of smoking cessation programs. These measures deal with behavioral, knowledge, and affective outcomes. *Behavior* measures focus on actual behaviors of program participants.

Knowledge measures are concerned with participant mastery of a defined set of information. *Affective* measures assess participants' attitudes and values.

Each measure is introduced by a brief description of the purpose of the assessment instrument, as well as procedures for administering, scoring, and analyzing the resulting data. All measures have been provided on detachable pages. At the beginning of Chapter Three, an overview description of the chapter's measures is provided to facilitate the selection of measures.

Local measure appraisal. Although the measures contained in this handbook have been created with considerable care and were pilot tested in small-scale tryouts, the measures have not yet been subjected to a formal empirical appraisal of their technical adequacy. Thus, in Chapter Four, a description is provided of how such technical appraisals of the handbook's measures can be carried out.

Annotated bibliography. Because evaluators and designers of smoking programs may wish to consult additional sources regarding program design and program evaluation, an annotated bibliography is provided in Appendix C to facilitate the handbook user's selection of such materials.

Amplified content descriptors. The information eligible for inclusion in the knowledge measures is provided in Appendix A as amplified content descriptors. Additional content that can be used for the generation of new items is also presented. However, these descriptors are not exhaustive accounts of smoking cessation content.

How to Use the Handbook

The particular ways in which the handbook is used will vary from setting to setting and from user to user. For instance, if a handbook user is relatively unfamiliar with the core notions in program evaluation, then a thorough reading of Chapter Two's treatment of program evaluation essentials is warranted. In addition, further reading based on the evaluation-related references included in the annotated bibliography would also seem useful.

For handbook users more familiar with program evaluation, primary attention will probably be focused on Chapter Three's measures. Although use of the measures will vary from situation to situation, a common four-step usage pattern is depicted in Figure 1.1.

Note that in Step 1, the measures are used to represent a range of potential program objectives. Clearly, an expanded range of options can lead to more appropriate decisions regarding what program objectives to pursue. In Step 2, after the measures for possible program evaluation have been reviewed, one or more measures are selected for use in the evaluation of the program. In Step 3, after the program evaluation measures have been selected, the program staff studies the measures intensively to discern if there are program design implications to be drawn from the measures. In Step 4, the measures are administered using one of the evaluative data-gathering designs described in Chapter Two and scored according to the scoring directions in Chapter Three. Finally, interpretations of the results are made.

It is important to remember that the handbook's measures are to be used for program evaluation, not individual decision making. Thus, if one of the handbook's affective

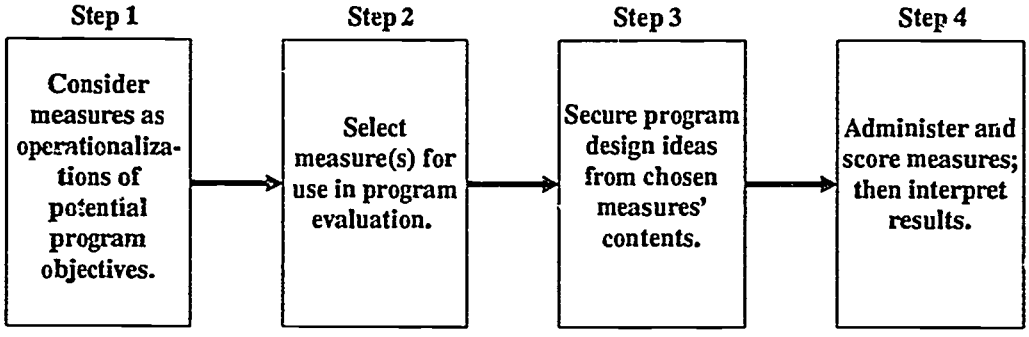


Figure 1.1: A four-step usage pattern of the handbook's measures

measures was used on a pretest-posttest basis, it is the *aggregation* of scores on the measure that provides us with an indication of the program's effectiveness. The measures were not designed to yield an accurate indication of an *individual* participant's status. Thus, it would be inappropriate to attempt to determine an individual participant's attitudes on the basis of the handbook's measures. The measures are relatively brief instruments designed to be administered without great intrusiveness. When the measures' scores are viewed in the aggregate, the measures can provide data of relevance to program evaluators. The data, however, should *not* be used for determining the status of individuals.

Another point related to use of the handbook's measures concerns the potential *reactivity* of certain measures, that is, the likelihood that if the measure is used *prior* to the program, the experience of completing a measure may cause participants to react differently to the program than had the measure not been administered. Reactivity is more frequently associated with affective measures rather than cognitive measures. Thus, handbook users will need to be alert to the possibility that a given measure, if administered prior to the program, will unduly sensitize participants to an aspect of the program.

To avoid such reactive effects, program personnel may need to divide participants into two subgroups so that only a portion of the participants receive any given potentially reactive measure. Such subgroups would not be given the same reactive measure both before and after the program. Rather, participants should be administered only post-program measures that they had *not* been given prior to the program. Indeed, two potentially reactive measures may be administered simultaneously under the conditions represented in Figure 1.2, where it can be seen that the pre-program performance of certain participants (one-half, for example) serves as a comparison for the post-program performance of other participants. Although a variety of data-gathering designs will be described in Chapter Two, the evaluator should employ care in using the handbook's measures so that they permit reasonable inferences regarding program effectiveness. Potential reactivity of measures should be examined when considering such designs.

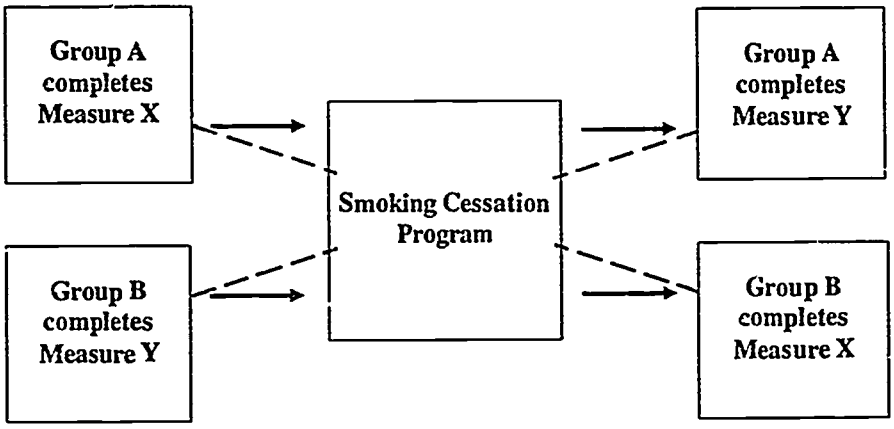


Figure 1.2: Using the handbook's measures to avoid reactive effects (Appropriate Comparisons = — — —)

Technical Quality of the Handbook's Measures

The measuring instruments to be found in Chapter Three were carefully constructed by an experienced test-development agency according to the guidance of prominent experts in the field of smoking cessation. All of Chapter Three's assessment devices were subjected to small-scale tryouts, revised on the basis of those tryouts, and reviewed by smoking cessation specialists.

At the outset of this handbook development project, it had been anticipated that all of the handbook's measuring instruments would be subjected to large-scale field tests so that substantial empirical evidence regarding the technical quality of the measures could be made available to handbook users. Unfortunately, that phase of the project could not be completed.

Thus, handbook users should be cautioned that, although the handbook's measures were developed with great care, there is currently no evidence available by which to ascertain the technical quality of the measures. Thus, handbook users must exercise caution in the use of Chapter Three's assessment instruments. In Chapter Four, as indicated earlier, a description is presented of the ways in which users of the handbook's measures, if they wish to do so, can carry out local studies regarding the technical quality of the measures that they find most suitable for their use.

Specific Smoking Cessation Concerns

This handbook is intended to help those who design and evaluate smoking programs. It is *not* intended to transmit content dealing specifically with smoking or with quitting smoking. For those readers who wish to acquire information about smoking, the list of sources located at the end of Appendix A contains introductory and advanced resources dealing with smoking *per se*. There are, however, a few issues related to smoking cessation programs that should be considered prior to a discussion of program evaluation.

The new cessation focus. Because smoking has been identified as one of the nation's most important public health issues, quitting smoking has become a national obsession. Since the United States Surgeon General's annual reports on the effects of smoking were first published in the 1960s, the American public has been searching for ways to modify, reduce, and eliminate the smoking habit. One recent Gallup poll suggested that two-thirds of all current smokers would like to quit. Of those who tried to quit, one-third went back to smoking within one week (Coelho, 1985). Most smokers know about the health risks of smoking, but most need something more than their own desire to get them to stop smoking.

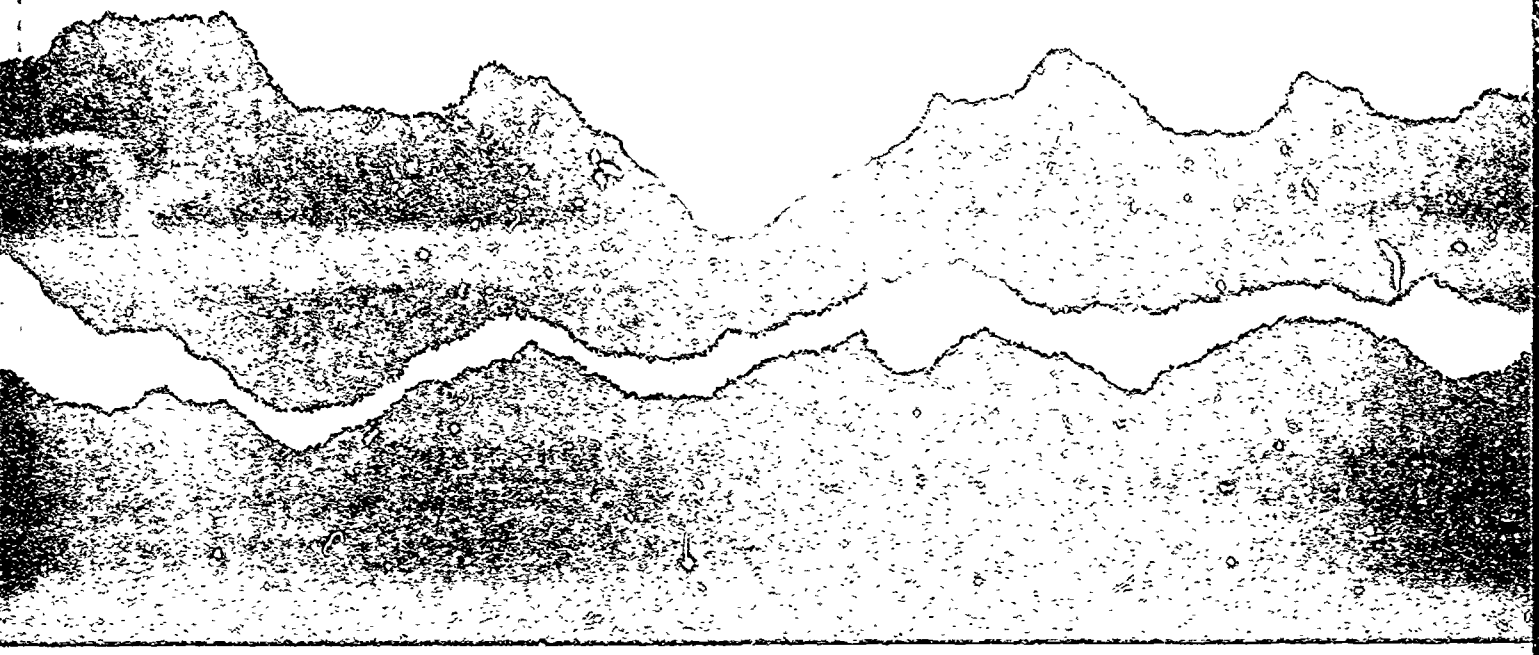
In addition to the consequence of injury to personal health, recent attention on the health risks of secondary smoke to nonsmokers has brought further pressure on smokers. The 1986 Surgeon General's report (USDHHS) notes that smoking programs need to be more concerned than ever about cessation of smoking rather than simple reduction because of the health risk to nonsmokers.

Both the strong internal desire to quit and pressures from society have motivated smokers to search for smoking cessation methods that work. In this handbook it is recognized that a wide variety of programs and methods can be effective for smoking reduction and cessation. The evaluation measures in Chapter Three have been designed to be applicable to an array of smoking programs. Use of this handbook in evaluating the effectiveness of such programs is the focus of the next chapter.

References

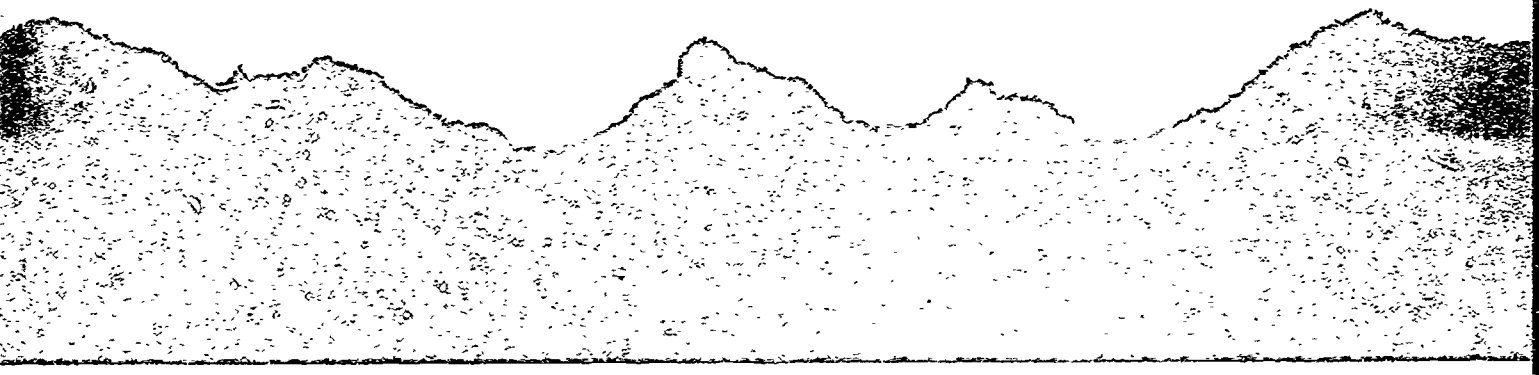
Coelho, R.J. (1985). *Quitting smoking: A psychological experiment using community research.* New York: Peter Lang.

U.S. Department of Health and Human Services. (1986). *The health consequences of involuntary smoking: A report of the surgeon general.* Washington, DC: U.S. Government Printing Office.



CHAPTER TWO

Essentials of Program Evaluation for Health Educators



Essentials of Program Evaluation for Health Educators

Education programs are intended to help people. Public school programs, for example, are intended to help youngsters acquire the skills and knowledge that they will need as adults. Similarly, health education programs are intended to promote participants' adoption of beneficial health-related behaviors. Yet, even though an education program might have been well intentioned, how do we know that the goals of the program were realized? Moreover, if a program is not meeting its goals, how can the program be made more effective?

Such questions constitute the core of program evaluation. In essence, evaluators want to discover whether a program has worked effectively and, if not, how it can be made more effective. When evaluation is used to improve programs, it can make a significant contribution to the well-being of program participants and, potentially, to the community at large.

In this chapter, the nature of program evaluation will be considered as it relates to health education programs. The following topics will be discussed:

- Focusing the Evaluation
- Rights of Participants
- Selecting Appropriate Measures
- When to Administer Measures
- Data-Gathering Design Options
- Sampling Considerations for Data Collection
- Data Analysis
- Reporting Results

The purpose of this chapter is not to promote a particular evaluation model for health education programs. Rather, the chapter deals with considerations central to any evaluation effort. It is hoped that evaluators* of smoking cessation programs will be able to apply the chapter's contents to their endeavors.

Focusing the Evaluation

The results of a program evaluation can be used to improve decisions about programs. Anyone setting out to evaluate a health education program, therefore, should focus the

---

* Sometimes a program evaluation will be conducted by an individual not affiliated with the program itself—an individual formally designated as a program evaluator. More frequently, however, an evaluation will be carried out by the personnel who are actually operating the program. Whenever the term "evaluator" is used in this handbook, it will refer both to the evaluator-specialist and to the program staff member serving as evaluator.

evaluation on the *decisions* that are likely to be made about the program, either while the program is being implemented or when it is concluded. In other words, if evaluators know what decisions are apt to be faced by those who will use the evaluation's results, then information bearing on those decisions should, if possible, be collected during the evaluation. To determine what these decisions are, an evaluator needs to have a clear understanding of the purpose of the program, the specifics of the program, and the individuals or groups who may use the evaluation's results. Focusing the evaluation involves considerations such as (a) the nature and role in the evaluation of program objectives, (b) the summative and formative functions of evaluation, (c) the cost of the program, (d) the extent to which observed changes in participants will also be attributed to the program, and (e) the extent to which program effects will be generalizable to other situations. Each of these considerations is discussed below.

Objectives and evaluation. Health education programs are designed to bring about worthwhile effects. Most health education programs, therefore, are organized around some form of program objectives that focus on such intended effects. In general, the more clearly these objectives are stated, the more useful they will be in carrying out an evaluation.

One way of conducting an evaluation is to determine the extent to which a program's objectives have been achieved. Program designers too frequently describe their objectives in such ambiguous, general ways, however, that it is impossible to tell whether such loosely defined objectives have been attained. It is for this reason that it can be beneficial for evaluators to work with program personnel, prior to program implementation, to create program objectives that clearly describe desired post-program participant behaviors.

Another potential pitfall when creating program objectives is the tendency to delineate a set of hyper-detailed objectives. Specificity does not automatically yield utility. Instead, decision makers can become overwhelmed by long lists of low-level, albeit behaviorally stated, objectives. For example, a program objective which states that participants be able to identify the heart as an organ affected by smoking is going to lead down a path toward numerous small-scope objectives. Recent thinking regarding instructional objectives suggests that program objectives, *while still measurable,* should focus on larger, more significant types of participant post-program behaviors. A more significant smoking-related objective, for example, might be that participants be able to identify the effects of smoking on the human body's organs. Today's health education programs, rather than being organized around 30 miniscule (and, therefore, potentially trivial) objectives, might better be focused on a half-dozen more general, but still measurable, program objectives.

Most evaluators agree, however, that there is substantially more to program evaluation than merely determining whether a program's objectives have been achieved. For example, there may be effects of the program that were not anticipated in the program's stated objectives. Evaluators need to be attentive not only to the effects of a program that were anticipated, but also to any unforeseen program effects.

Summative and formative functions. Summative evaluation addresses the question of whether a program, in its complete and final form, is effective. The decisions associated with the summative evaluation are essentially go/no-go decisions, such as whether to continue a health education program or, perhaps, whether to disseminate the program more widely.

Formative evaluation addresses questions associated with improving a program that is "under development," that is, still modifiable. The decisions associated with formative evaluation focus on ways to improve particular parts of the program. Formative evaluation is an ongoing endeavor conducted as the program is designed, installed, and maintained. Whereas summative evaluation's mission is to provide a final judgment about a program's overall merit, formative evaluation's mission is to bolster a program's quality on a continuing basis. The effective formative evaluator functions less as an external judge and more as a collaborating member of the program team. The formative evaluator's task is to monitor the program so that it can be improved.

Almost all programs are, at least to some degree, modifiable. Hence, only in rare cases do evaluators appraise a health education program in its complete and final form. One such instance might involve a packaged, manual approach to smoking cessation. For example, if the program were found to be effective via a summative evaluation, a commercial publisher would distribute the manuals nationally. In most cases, however, health education programs can be modified and improved. Thus, a formative, improvement-oriented evaluation can be carried out for most health education programs.

Cost-analysis considerations. Program evaluators are often so concerned about detecting the effects of programs that they fail to consider the *costs* of those effects. Yet decision makers need information regarding not only the effects of a program, but also the resources required to achieve those results. For this reason, program evaluators should carefully isolate and communicate the relative costs of programs. For example, information should be collected that can show how much Program A costs to produce a given result compared to the cost of Program B to produce a comparable result. Judgments about a program's impact without considerations regarding its costs are potentially superficial. In recent years, there has been much attention to cost-analysis strategies. Although consideration of those procedures is beyond the scope of this handbook, serious evaluators of health education programs would do well to delve more deeply into cost-analysis procedures.*

Attributing observed changes to the program. Characteristically, an evaluation seeks to determine whether individuals have changed as a result of their participation in a program. The key issue is whether pre-program to post-program changes in the status of participants are attributable to the program itself or to other extraneous factors. Examples of extraneous factors are participants' maturation, their familiarity with the measures used in the evaluation, or their reactions to nonprogram events such as a health-related, mass media campaign. This issue revolves around the evaluator's ability to properly infer that the program itself caused any observed changes in participants. Technically, the degree to which evaluators can validly infer that a program caused a set of observed changes is referred to as the *internal validity* of the evaluation study. Ideally, an evaluation's data-gathering design should help to rule out explanations other than the program itself for observed changes. (Data-gathering design options are discussed later in this chapter.) If evaluators are unable to attribute observed changes to the program, they will have difficulty in determining program quality.

---

* For additional information about cost-analysis approaches, see Annotated Bibliography Nos. 1, 28, and 29.

Generalizing program effects. A related issue is the extent to which the findings of an evaluation study can be generalized to other situations. The issue here is whether the program would be expected to produce similar results with, for example, a different group of participants, slight variations in the program, or changes in program personnel. The degree to which the results of an evaluation study can be generalized elsewhere is technically described as the study's *external validity.*

If evaluations are generalizable, they can provide useful information to (a) program personnel regarding the range of conditions under which the program is effective and (b) other health educators who may wish to adopt an already "evaluated" health education program. A smoking cessation program that works well in one setting may provide helpful guidelines for those wishing to operate other smoking programs. Typically, however, a local evaluation should be conducted once the program has been adopted.

It is important to distinguish between a program's causative power and the program's generalizability, because different information may be required to establish each factor. Procedures that limit the number of extraneous variables in the evaluation (e.g., including only males) increase internal validity but, at the same time, limit generalizability. Evaluators must try to balance the problems associated with threats to internal and external validity by selecting a data-gathering design that best addresses the information needs of program personnel as well as of those external to the program who may be interested in adopting the program elsewhere.*

Rights of Participants

Health education programs are designed to improve individuals' health and well-being. When such programs are evaluated, therefore, the focus is typically on a program's impact on human beings. Some evaluators, however, become so caught up with the importance of appraising a health education program that they overlook the rights of the individuals who take part in the evaluation. Two important rights are those of informed consent and confidentiality.

Informed consent. Evaluators, just as researchers, should be guided by a profound respect for human dignity. Therefore, they should not engage in evaluative activities that in any way demean participants. Prominent among the considerations that should guide evaluators is the concept of *informed consent.* Informed consent requires that an evaluator secure, in advance of the study, permission from the participants in an investigation to gather data from them. This consent is obtained *after* the potential participants have learned about the nature of the investigation and what their role would be, because that information may influence their decision to participate. Informed consent eliminates the possibility of making individuals unknowingly serve as subjects in an evaluation.

Two different approaches to securing informed consent have been employed by program evaluators. The first of these, *active informed consent,* obliges an evaluator to obtain, in writing, a statement from each participant indicating that the individual is willing to

---

* For additional information about internal and external validity issues, see Annotated Bibliography Nos. 8, 11, 12, and 16.

participate in the evaluation. The significant aspects of the evaluation must be described in the written permission form so that potential participants are fully informed when they give their consent.

An evaluator using the second approach, *passive informed consent*, supplies descriptions of the evaluation's essentials to all program participants and provides them an opportunity to register, in writing, their unwillingness to participate in the study. In other words, when a passive informed consent approach is used, participants return the forms supplied to them only if they are *not* willing to participate in the evaluation study. Of the two approaches, the active informed consent strategy typically results in fewer participants because those individuals who do not provide consent forms must be excluded from the study. Because evaluators who conduct studies involving school-age children are obliged to secure informed consent from underage participants' parents or guardians, a passive informed consent strategy is often adopted due to the difficulty of securing active informed consent from individuals who are not participating in the program themselves.

Procedures for developing forms for both of these approaches to securing informed consent are described in Appendix B. The actual forms to be used in an evaluation would need to be created so that they are specifically relevant to the program involved.

Confidentiality. Another consideration when dealing with human subjects is the *confidentiality* of all information gathered during an evaluation. Because the evaluator is not concerned with an appraisal of individual participants but, rather, with gauging the effectiveness of a health education program, ensuring participant confidentiality usually poses no problem. Evaluators must, however, devise protective safeguards, such as anonymous completion of forms and careful handling of data, to ensure both the appearance and reality of confidentiality.*

Selecting Appropriate Measures

Although there are various approaches to program evaluation, almost all share one common feature, namely, the systematic gathering of evidence regarding a program's effects. To secure evidence of program effects, evaluators usually employ measurement instruments. Some instruments, however, are far more suitable for assessing a program's effects than others.

Criterion-referenced measurement. For more than two decades, educational measurement specialists have directed increasing attention toward an emerging form of assessment known as criterion-referenced measurement. In comparison to norm-referenced measurement, which attempts to ascertain an examinee's status in relation to the status of other examinees, criterion-referenced measurement attempts to ascertain an examinee's status in relation to a clearly defined set of behaviors. The essence of a criterion-referenced instrument is the clarity with which its accompanying descriptive materials explain what is being measured. Because norm-referenced instruments emphasize *relative* comparisons among examinees, they often do not provide a clear description of exactly what it is they are assessing. In

* For additional information about the rights of human subjects and the ethics of evaluation, see Annotated Bibliography Nos. 2, 26, and 38.

contrast, criterion-referenced instruments are *absolute* measures, designed to determine exactly what it is that examinees can or cannot do, without reference to the performance of other examinees. Thus, criterion-referenced tests provide a clearer description of what they are measuring.

It is the clarity regarding what is being assessed that renders criterion-referenced measures ideal for the evaluation of health education programs. Consistent with the mission of providing useful information for decision makers, criterion-referenced instruments describe the precise nature of what is being measured. Hence, when criterion-referenced measures are used to gather evidence in program evaluations, decision makers can accurately interpret the evidence being supplied.*

Attributes of well-constructed measures. All instruments, whether norm-referenced or criterion-referenced, should measure what they are measuring with consistency. The consistency with which an instrument measures is known as its *reliability.*** There are several different indices that can be computed to reflect an instrument's reliability. The kind of reliability data needed to appraise a measure for possible use in an evaluation study should be consonant with the way the measure will be used in that study. If a measure is to be used on a test-retest basis, for example, then information about that type of reliability is germane. If alternate forms of a test are to be used, for instance, in a pretest-posttest situation, then evidence should be available regarding alternate-forms reliability so that the evaluator can determine whether or not the two different forms are sufficiently equivalent.

It should be noted that when a health education program is being evaluated, attention should be directed to the impact of the program on a *group* of participants. Thus, the consistency to be sought when measurement instruments are used for program evaluation is consistency for a group of participants' scores. When dealing with individual participants, the measures must yield *individual* or diagnostic consistency.

A second critical attribute of a properly constructed measure is that it yields scores from which valid inferences can be drawn. An instrument is often said to be valid "if it measures what it purports to measure." Such a statement, however, is technically in error. Tests themselves are never valid or invalid. Rather, it is the *interpretations* made from test scores that are valid or invalid.

There are several types of validity evidence, each yielding somewhat different but conceptually related indications about our ability to make valid inferences from a measure. Evidence of validity is, in the opinion of most measurement specialists, the most important consideration in judging the adequacy of measurement instruments. Program evaluators should make sure they are knowledgeable about methods of securing validity evidence.***

---

* For additional information about the nature and development of criterion-referenced measures, see Annotated Bibliography Nos. 7, 24, and 34.

** For information about determining the reliability of measurement instruments, see Annotated Bibliography Nos. 3, 18, 19, 23, 27, and 34.

*** For information about obtaining validity evidence regarding measuring instruments, see Annotated Bibliography Nos. 3, 18, 19, 23, 27, and 34.

A final consideration in appraising the quality of measures used for program evaluation deals with the presence of *bias* in the assessment devices. During the past decade, measurement specialists have become particularly aware that many educational assessment devices contain items biased against particular subgroups, such as ethnic minorities or women. An example of a biased test item would be a knowledge question that, because of peculiarities in its content or wording, is more difficult for women to understand and answer correctly than it is for men, even though the men and women have an equivalent amount of knowledge regarding the particular concept being tested.

Another type of bias that can adversely influence examinee performance arises when test items are offensive to particular groups of individuals. For example, if a test item includes content that is seen to be derisive to members of particular ethnic groups, then examinees from those groups are not apt to perform at their best on the item. Their warranted agitation over the offensive content is likely to interfere with their responses to that item as well as to subsequent items.

There are now available both judgmental and empirical techniques for detecting the presence of biased items. These approaches should be used to identify, then eradicate, bias in a measure's items.*

Finally, it is important to note that any given instrument may not possess all of the qualities discussed above. Often evaluators must choose among measures that embody some but not all of the elements described here, that is, (a) descriptive clarity, (b) reliability, (c) validity, and (d) absence of bias. Another important point is that merely because a measure is *labeled* in a particular way, for example, as criterion-referenced or as nonbiased, that does not automatically indicate that it is of sufficient quality to be used in evaluating a health education program. Scrutiny of all aspects of the measure's quality is requisite.

When to Administer Measures

Decisions regarding when to administer measures depend on the data-gathering design selected. Conceivably, there are four temporal periods during which it may be useful to obtain evaluative information about participants of health education programs. There may also be reasons for repeated measurement during some of these periods. These periods are depicted in Figure 2.1.

Pretests. Often it is useful to have information about participants prior to their starting the program. Such information, typically referred to as pretest data, may be used to identify participant needs so that instruction can be targeted directly at those areas. In addition, pretest data can be compared with data collected at the end of a program. Such a comparison can provide a measure of program impact.

En route tests. Measures can also be administered during a program to secure current readings on the status of participants. For purposes of formative evaluation, en route data can be used to redirect resources during the program by providing program personnel with ongoing status-checks on participants' progress. Thus, en route tests may be even more

* For information about methods for avoiding test bias, see Annotated Bibliography Nos. 6 and 33.

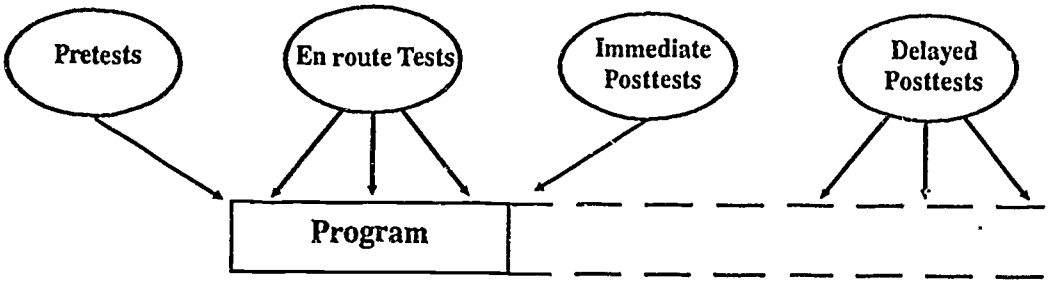


Figure 2.1: Possible measurement times in program evaluation studies

useful than tests administered at the end of the program, because en route measurement provides information while there is still time for program personnel to act on it. This type of assessment is most appropriate for programs of long duration (e.g., several months or more).

Immediate posttests. Measures are commonly administered following a program. The data from posttests can be compared with pretest data to examine changes in participants from the beginning to the end of the program. Participants' posttest performance can also be contrasted with posttest scores from participants in other programs. In addition, posttest data provide an indication of the absolute status of participants on the variables of interest at the completion of the program.

Delayed posttests. Data from delayed or follow-up posttests are often as important or more important than immediate posttest data in evaluating a health education program. Delayed posttest data might be secured, for example, several months after a program's conclusion. Far too frequently data collection efforts are limited to those times when measurement is most convenient. Ultimately, however, health educators should be interested in effecting long-term, rather than short-term, behavioral, affective, and cognitive changes. It is nearly impossible to infer such long-term changes on the basis of information gathered solely at the end of a program. The long-term objective of quitting smoking for a lifetime is the desired result of most smoking programs rather than the short-term goal of more knowledge of smoking effects. For most health education programs, some follow-up measurement is usually warranted.

Clearly, it is not sensible to administer all measures at all time periods. Evaluators, in collaboration with program personnel and other interested parties, need to select a measurement scheme that focuses on the most appropriate times for gathering data. Just as it is desirable to avoid administering an excessive number of different measures, it is also necessary to avoid an excessive number of administrations. It may be useful to administer certain measures (for example, a brief behavioral self-report measure) on a continuing basis; other more time-consuming measures might be administered less frequently. Decisions about when to administer measures should be guided by common sense, attentiveness to participants' feelings, the efficient use of resources, and any conventional expectations, such as when a delayed posttest is ordinarily given.

Data-Gathering Design Options

It is sometimes thought that program evaluations must include complicated and elaborate data-gathering designs in order to yield decisive and compelling data. This is simply not the case. Program personnel and evaluators should try to conduct evaluation studies and gather data in such a way that the ambiguity of results can be reduced to a minimum. That is, evaluations must attempt to determine whether a program works and what makes it work or what prevents it from working. Data-gathering designs serve as the means to this end by setting forth the procedures to be used in exploring the nature and impact of a program.

The data-gathering design that an evaluator chooses for an evaluation will determine the inferences the evaluator can make about a program's overall impact on participants and the effectiveness of its various components. To select the best designs for evaluation studies, evaluators must have a broad knowledge of the available data-gathering design alternatives and the strengths and weaknesses associated with each. Evaluators must also work closely with program staff to determine what decisions are at issue regarding the program. No evaluation study will be perfect; every evaluation leaves some questions unanswered. Evaluators need to be clear regarding what they have learned about a program and the degree of certainty associated with their findings, and they must convey this information to appropriate audiences.

An important concept related to data-gathering designs is randomization. Randomized selection and assignment are described below, followed by brief descriptions of the most common data-gathering designs available for evaluators of health education programs.

Randomization. One technique that can prove useful to evaluators is *randomization,* which involves the selection or assignment of participants in a nonsystematic manner, such as by using a table of random numbers (found in most statistics texts). A prominent application of randomization in program evaluation is *randomized selection* of subjects. This sort of randomization is particularly important when the evaluator wishes to generalize from the results of a study to a larger population. When the participants taking part in the program to be evaluated have been selected at random from a larger population of potential participants, then the evaluator can be reasonably confident that those involved in the evaluation will be representative of that larger population. There is less likelihood that the participants being studied in the evaluation are atypical, which would make it inappropriate to generalize the evaluation's results to the population at large. Randomized selection of subjects may also be useful when there are more applicants than vacancies for a program.

Another use of randomization is to assign participants to different "treatments" or programs. If an evaluator wishes to compare the effects of different treatments, then the evaluator wants the participants in each treatment to be as equivalent as possible. To this end, evaluators can employ a *randomized assignment* procedure whereby individuals are randomly placed in the treatments or programs to be compared.

The two procedures of randomized selection and randomized assignment are illustrated in Figure 2.2. Note that participants are randomly selected from the pool of potential participants, and then randomly assigned to either Program A or Program B.

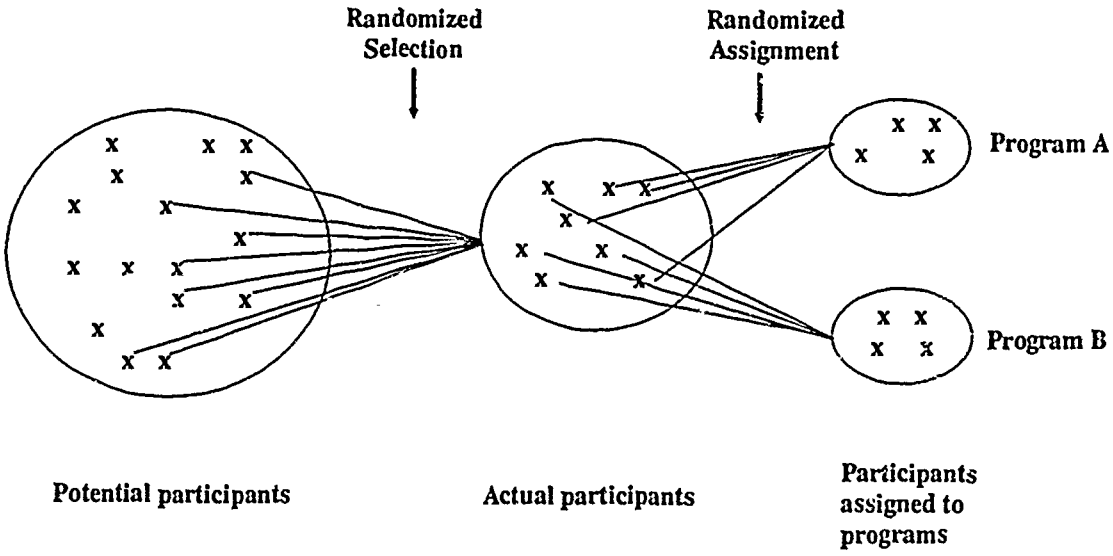


Figure 2.2: Randomized selection of participants from pool of potential participants and randomized assignment of participants to programs

The use of randomization techniques does not necessarily create equivalent groups. For example, if an evaluator were to randomly assign 50 potential participants in a company's smoking cessation program to treatment and no-treatment groups, it is still possible that one of the groups would contain individuals who, when pretested, were significantly different in some important aspect from those in the other group. In such instances, evaluators must rely on statistical procedures in an effort to compensate for such disparities. In most cases, however, use of randomization will create groups of sufficient equivalence that such statistical adjustments are not needed.

In practice, program personnel often may not have the luxury of constituting groups via randomized selection or assignment. For example, local school board policies might require that *all* youngsters be provided with any program regarded as potentially beneficial. When randomization is not used, it is especially important to collect and examine descriptive data about participants to determine where pre-program group differences occur and to consider the ways in which such differences may influence post-program data. Even if randomization is impossible, attempts to constitute comparison groups with individuals as equivalent as possible can help minimize the influence of preexisting participant differences.*

Seven different data-gathering designs of potential utility for evaluators of health education programs will be presented below. Each data-gathering design will be described and depicted schematically. Some of the major factors involved in the selection of data-gathering designs will be addressed.

---

* For additional information about randomization, see Annotated Bibliography Nos. 8 and 25.

The case-study design. Consider a six-month health education program aimed at modifying participants' knowledge about the effects of smoking on health. If participants' knowledge were measured only at the close of the program, we could describe the data-gathering approach as a *case-study design* and represent it schematically as shown in Figure 2.3.

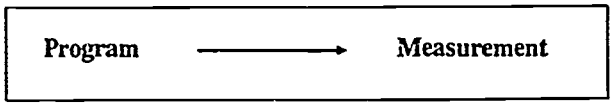


Figure 2.3: Case-Study Design

If this were the design employed in an evaluation, what could an evaluator tell about the program's impact on participants' knowledge? How confident would an evaluator be that participants' knowledge about the effects of smoking was attributable to the program?

It would be difficult, with confidence, to attribute any effects to the health education program. The program, indeed, may have been totally ineffectual. In fact, participants' post-program knowledge might be identical to their knowledge before the program. The participants could be demonstrating knowledge that they brought to the program, not that they acquired during the program. Because we have no measure of participant knowledge prior to the program, we cannot distinguish between preexisting knowledge and knowledge acquired as a result of the program. Hence, with the case-study design, it may be impossible to determine whether the program had any impact on participants.

Even though attributions of causality are often unwarranted, it may be possible to secure useful program evaluation data with such a data-gathering design. Suppose, for example, that a health education program is promoting a body of knowledge so advanced that few, if any, individuals would be familiar with it. In such a setting, one could assume that participants' post-program knowledge is attributable to the program's impact because participants would almost certainly not have acquired the knowledge without the program. It might not be worth the resources necessary to implement a data-gathering design capable of conclusively demonstrating that participants began the program unfamiliar with the knowledge being promoted.

This example illustrates an important data-gathering consideration, namely, that the chief mission of data-gathering designs is to *rule out plausible rival explanations,* that is, explanations other than the program's impact that might account for the post-program status of participants. If there is reason to believe that participants' pre-program status may account for their post-program status, then a data-gathering design should be selected that permits the evaluator to rule out this rival explanation.

The one-group pretest-posttest design. Now suppose that, to avoid the major shortcoming of the case-study design, an evaluator measures participants' behavior both before and after a health education program. This data-gathering approach can be described as a *one-group pretest-posttest design* and can be represented as shown in Figure 2.4.

Assume an evaluator uses the one-group pretest-posttest design and that the data reveal a substantial shift toward more desirable behaviors between the initial and the final

Figure 2.4: One-Group Pretest-Posttest Design

measurement. Can this change in behaviors be ascribed to the program? Unfortunately, the evaluator cannot be sure. There are many other factors, totally unrelated to the program, that may have influenced participants' behaviors. For instance, if a smoking cessation program emphasized the relationship between smoking and cancer, and at the same time a number of prominent people died from smoking related cancers, such news may have influenced participants' views regarding smoking and cancer. Evaluators of programs that serve children must also consider the possible effects of maturation during the time the program is offered. Participants' increased maturity may cause pre-program to post-program shifts in behaviors. The program itself may have contributed nothing to the measured shift of behaviors. Such extraneous factors decrease the evaluator's ability to draw defensible conclusions about the program's impact.

As was true with the case-study design, however, if there are no plausible rival explanations for the posttest results, the one-group pretest-posttest design can be suitable for the task at hand. In fact, this simple yet serviceable design is often used in formative evaluation.

The one-group pretest-posttest design requires measurement before as well as after a program. This points to a commonly accepted but often overlooked principle of effective program evaluation. Evaluation is most effective when it is initiated at the beginning of a program. If evaluators are not called in until the end of a program, they may be hampered in their efforts to design a credible program evaluation.

The nonequivalent control/comparison group design. Program evaluators can eliminate some of the more common rival explanations for changes in participants' behaviors by using data-gathering designs in which either comparison or control groups are employed. The use of a control group (untreated individuals) or a comparison group (individuals receiving a different program) requires two groups that are assumed to be relatively similar (before the program) on all related variables. When using these designs, the evaluator should attempt to secure two groups that are as similar as possible. Because the two groups are not randomly assigned to the two conditions, however, they cannot be assumed to be *equivalent*, hence the design's designation as a "nonequivalent" control or comparison group design.

In the control-group version of this design, only one of the groups is given the program to be evaluated; the other group is left untreated. This data-gathering design, known as the *nonequivalent control group design,* is illustrated in Figure 2.5.

In this design, a control group (Group 2) is assessed before and after the program, but it never receives the program itself. Assuming that the groups were similar before the program, if the program participants' behaviors change while the behaviors of those in the

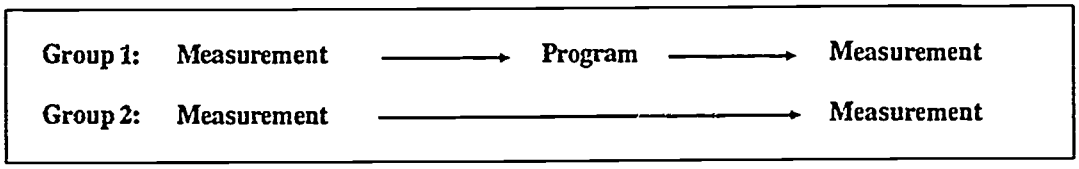


Figure 2.5: Nonequivalent Control Group Design

control group remain the same, the evaluator can be reasonably confident that the program caused the change.

The use of an *untreated* control group may strike some health educators as a particularly unsavory data-gathering ploy. After all, health educators design their programs to benefit participants. To withhold such programs from individuals, even for the important purpose of evaluating the program's effectiveness, seems downright reprehensible. Yet, the individuals from whom the program is withheld, that is, the members of the control group, can be given the program *subsequently,* as soon as the evaluation study has been concluded. Also, in some situations there are more program applicants than can be accommodated, and, therefore, some prospective participants must be denied access to this program under any circumstances. Those who are not admitted to the program could be used as a control group, and admitted to the program the next time it is offered.

A variation of the nonequivalent control group design involves the use of a comparison group, that is, a group receiving a different program or a different treatment. Program evaluators frequently find themselves studying the quality of two or more competing programs. Thus, the evaluator focuses on the relative virtues of two or more different programs rather than on a contrast between a single program and an untreated control group. A schematic depiction of a *nonequivalent comparison group design,* in this instance contrasting two different programs, is presented in Figure 2.6. As indicated above, more than two groups can be employed when using a nonequivalent comparison group design. An evaluator using this design can be fairly certain that, if the groups were similar before the program, any differences in post-program behaviors are due to the differential impact of the two programs.

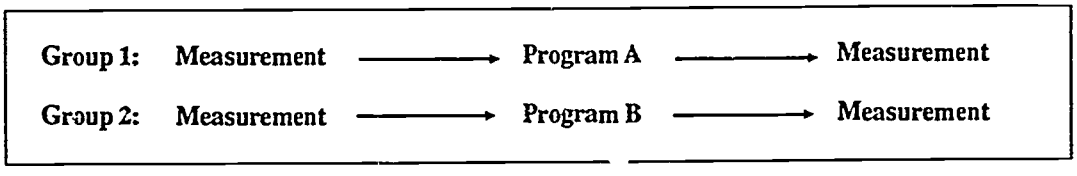


Figure 2.6: Nonequivalent Comparison Group Design

There are, however, potential problems with the nonequivalent control/comparison group designs. It may be that the initial measurement was *reactive*. A reactive measurement is one that, by itself or in combination with the program, influences participants' behavior. Attitude inventories and self-report questionnaires about behavioral practices are

notoriously reactive. For example, a questionnaire administered before the program might alert participants to the importance of a desired behavior. This would heighten their attentiveness when the program dealt with content related to that behavior and, as a consequence, influence their performance on the second measurement.

Moreover, measurement is expensive. Measuring the status of control groups requires valuable evaluation resources. Time and money can often be better spent studying the program being evaluated rather than studying a no-treatment control group of little intrinsic interest. Health educators should not ritualistically employ control groups in their designs if the questions at issue can be answered without the use of untreated groups.

The pretest-posttest control/comparison group design. There are two data-gathering designs that are of particular value to program evaluators if randomized assignment is possible. The first of these is the *pretest-posttest control group design*, illustrated in Figure 2.7.

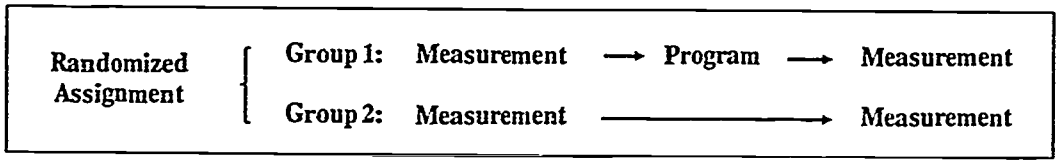


Figure 2.7: Pretest-Posttest Control Group Design

The difference between this design and the previously considered nonequivalent control group design is, of course, the randomized assignment of subjects to the two groups. This feature of the design is a particularly important one, because creation of two or more groups using randomized assignment is an effective way of promoting equivalence between the groups, especially if the number of subjects in each group is large (say, 30 or more). Equivalence of groups at the beginning of the program strengthens the inference that any differences at the conclusion of the program are due to program impact.

By using comparison groups, that is, two or more program groups, instead of an untreated control group, the evaluator would be using a *pretest-posttest comparison group design*, shown in Figure 2.8.

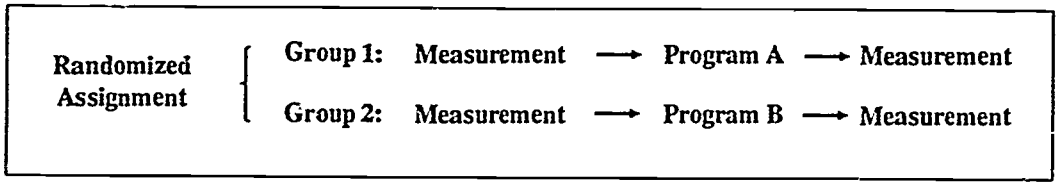


Figure 2.8: Pretest-Posttest Comparison Group Design

Because pretests are used in both of these designs, the possibility of reactive preprogram measures is still present. For situations in which reactivity is of great concern, a different data-gathering design, described next, has much appeal.

The posttest-only control group design. In situations where a measure is likely to be reactive, the evaluator can rely on a clever data-gathering design that effectively dodges the reactivity problem. This *posttest-only control group design* is depicted in Figure 2.9. This design is the same as the pretest-posttest control group design, except that there is no pretest.

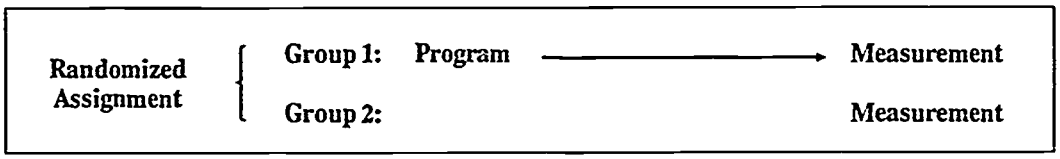


Figure 2.9: Posttest-Only Control Group Design

In this design, neither Group 1 nor Group 2 is pretested, but because of random assignment the groups can be considered equivalent prior to Group 1 receiving the program. Not pretesting Group 1 effectively avoids a pretest's potentially reactive effect on program participants. To assess the impact of the program, it is possible to contrast the *posttest* performances of Groups 1 and 2. As with the other control group designs, the untreated control group could be given the program the next time it is offered.

The basic dividend of the posttest-only control group design is that by measuring an untreated, randomly assigned control group, the evaluator secures an estimate of how program participants would have responded on a pretest, but without introducing the potentially reactive effects of a pretest. Although the diagram for this design suggests that the measurements be made for both groups at the conclusion of the program, it is possible to measure the untreated control group earlier if that seems advisable.

Multiple measures over time. There are certain situations in which health educators may wish to appraise the effects of their programs on the basis of periodic measurements, for example, by using regularly administered questionnaires or data that are routinely recorded. For instance, suppose when evaluating a "supervisor's smoking cessation" program, the evaluator was interested in the number of smoking-related referrals a company's supervisors make for their employees. Assuming that such information is available from the firm's health records, the evaluator might study records at periodic intervals before, during, and after the program. By observing the frequency of referrals during different time intervals, the evaluator would have valuable information regarding program effects.

A number of the most commonly used data-gathering designs have been described. There are other, more complex designs than those treated here.* Complexity, however, is rarely an asset if a simpler, more straightforward design is appropriate.

---

* For additional information about evaluation design options, see Annotated Bibliography Nos. 8, 11, 22, 23, and 35.

Sampling Considerations for Data Collection

The data-gathering requirements of an evaluation can become a burdensome intrusion into an ongoing health education program. Participants in a smoking cessation program can become more than mildly anxious if evaluators are requiring them to complete measures every hour or so. Accordingly, evaluators should conduct their data-gathering activities in the least intrusive manner possible. One way to minimize an evaluation's intrusiveness is by relying on sampling techniques, such as person-sampling and item-sampling, each of which is described below.

Person-sampling. To estimate how a large group of people would respond on a particular measure, it is not necessary to administer the measure to all the individuals in the group. Instead, a smaller group can be selected. This smaller group can be either a *simple random sample* or a *stratified random sample,* that is, a sample stratified on the basis of program-relevant factors such as age, sex, and socioeconomic status. Assuming that the sample is randomly selected, the evaluator can estimate the status of the total group based on the responses of the sample.

Suppose, for example, that the evaluator wants to use a measure to determine participants' perceived ability to refrain from smoking. Assuming that there is a reasonably large number of program participants, say 50 or so, the evaluator could randomly select half of the participants and administer the measure to this group only. In essence, this approach allows the evaluator to infer how the total group of participants would score on the measure, even though only half of the participants completed it. Thus, it is possible to estimate total group performance with only half the amount of participant time required for data gathering.

Using a similar sampling procedure, evaluators can administer two or more measures at once in the time it takes to administer one. Suppose that two measures are to be given to program participants. The evaluator can randomly assign one measure to half of the participants and the other measure to the remaining participants. Each participant needs to respond to only one measure, but the evaluator can derive defensible estimates of how all the participants would have responded on both instruments.

Item-sampling. In addition to sampling persons, as in the previous examples, it is also possible to sample items, so that different sets of items from a program evaluation measure are randomly selected to be administered to different persons. Using this approach, the evaluator gives each participant only a sample of the items on any particular measure. For example, suppose a program evaluator wishes to administer a 30-item test. Given 60 participants in the program, the evaluator could divide the test into three sets of 10 items each and administer each set of 10 items to 20 different participants. In this way, the total group's performance on the whole test can be estimated. This approach to data-gathering requires only one-third of the time that would have been required to administer the total 30-item test to all participants.

Sample size. Given the relatively small number of participants in some health education programs, is it really appropriate to sample either persons or items? How large must groups be before these sampling procedures can be sensibly used? Unequivocal answers to these questions do not exist. Some texts on sampling provide rules of thumb for estimating the

size of samples needed for detecting group differences in relation to the magnitude of differences sought and the nature of the groups being sampled. At best, though, these rules provide only rough estimates. It is important to recognize that the task of identifying a sufficiently large sample is more difficult than usually thought.

The variability of participants' anticipated performance on the measures is the primary determiner of the sample size necessary. If it is expected that participants' scores on a test will be relatively homogeneous, a smaller number of respondents will be needed than if participants' scores are expected to vary widely. Thus, if on a measure of knowledge about the effects of smoking on society, for example, some of the participants are expected to know many effects and others are expected to know very few, reasonably large numbers of participants (e.g., 20) should respond to any one item.

Intuitively, one recognizes that when working with a very small group of program participants, the use of these sampling techniques is risky. For instance, if there were only 15 participants in a program, few evaluators would try to split these participants into three groups of five each for purposes of taking different sets of items. Even though each group represents one-third of the total population, there is too much likelihood that a sample of five individuals would not properly represent the total group. One or two atypical participants in a five-person group would render the group's average performance unrepresentative of how the larger group would have performed.

It should be noted that when employing procedures such as person-sampling or item-sampling, an evaluator is focusing on a group of participants *in the aggregate*. Because evaluations are typically concerned with the effects of programs on groups of participants, the use of sampling procedures is usually appropriate. If, however, program personnel need individual data on all examinees, then sampling should obviously not be employed.*

Data Analysis

A frequent question asked of an evaluator is whether a study's results are statistically significant. For example, could the observed changes in program participants' knowledge or behavior from pretest to posttest have occurred simply by chance? Statistical tests are used to answer this type of question. Consideration of statistical analysis procedures, however, is beyond the scope of this handbook. Indeed, for those genuinely unfamiliar with statistical analyses, attempts to boil down such a complex subject into a few pages would be unwise. Thus, just a few comments will be made here regarding data analysis. Because there are many subtle choice-points in the statistical analysis of evaluation data, evaluators who are not well versed in at least the more common statistical procedures should probably enlist the aid of someone who is.

There are two basic classes of statistics, namely, descriptive statistics, such as the mean, and inferential statistics, such as the *t* test. *Descriptive statistics* help evaluators portray a group's performance on a given measure. For example, an evaluator might describe a set of participants' scores via the mean score (the scores' central tendency) and standard deviation of the scores (the scores' variability). Because the mean and standard deviation are

* For additional information about sampling procedures, see Annotated Bibliography Nos. 9 and 10.

frequently used, program evaluators should know how to calculate and interpret them. Any introductory statistics book for the social sciences will serve as a reference for this information. *Inferential statistics* help evaluators determine whether an observed difference between pre-program and post-program scores is *statistically significant,* that is, whether such a difference could have occurred because of chance alone. If the probability is small that the results are due to chance, the evaluator can, with reasonable confidence, attribute the results to the program.

Statistical significance, however, does not imply *practical significance.* A small difference between the average scores of two groups can be statistically significant, particularly when large numbers of participants are involved, yet be of no practical consequence whatsoever. Health educators will need to make sensible determinations regarding whether the magnitude of an observed difference, even though statistically significant, is sufficiently important to warrant action. In other words, although evaluators of health education programs should often carry out statistical significance tests, they should not be unduly swayed by the results of such analyses. Common sense must always be applied in interpreting the meaning of a statistically significant result.*

Reporting Results

Reporting the results of an evaluation study is a more difficult undertaking than is usually recognized. Considerable attention must be given to the procedures employed to report the results of health education program evaluations. When reporting evaluation results, as when focusing and planning the evaluation, the evaluator must be responsive to the needs of program decision makers. A few key considerations should be kept in mind when reporting evaluation results.

Evaluators must report their results to decision makers in a timely fashion. It does no good to deliver an evaluation report several weeks after key program decisions had to be made. Evaluators must also be careful to disseminate their findings to all appropriate audiences If possible, an evaluator should circulate the preliminary draft of a program evaluation report to program personnel so that they can react to its accuracy and objectivity.

The decision makers whom evaluators are assisting may have scant experience with quantitative data. As a consequence, complicated statistical presentations may be of little value to them. Evaluators should select data presentation procedures that will match the technical sophistication of the decision makers involved. In any evaluation report, there is nothing wrong with simple graphs or "percentage correct" tables. The more intuitively comprehensible the data presentation techniques, the better they are. Program evaluators should provide straightforward presentations of data without fearing that such approaches will be regarded as too elementary. Adequate technical back-up can be appended as necessary to the final report.

Evaluators should not be reluctant to make speculations based upon their knowledge about a program, but these conjectures should be identified as such. Similarly, if any of the

---

* For additional information about data analysis, see Annotated Bibliography Nos. 25, 36, 39, 43, and 45.

evaluation's findings are equivocal, the evaluator should inform concerned audiences of this fact. Honesty and objectivity are the hallmarks of effective evaluation reporting.

In addition, because decision makers are typically busy people, evaluators should strive for reasonable brevity in their reports. The preparation of executive summaries to accompany lengthy reports is a useful practice. Voluminous evaluation reports are almost certainly destined to go unread. Terse, easily read reports are much more likely to make an impact on decision makers.

The whole thrust of the evaluation enterprise is to facilitate better decisions. Decision making will *not* be illuminated by complex, lengthy, or otherwise incomprehensible presentations of evaluation results. The quality of decision making can be enhanced only if an evaluation's results are reported in a way that can be clearly understood.*

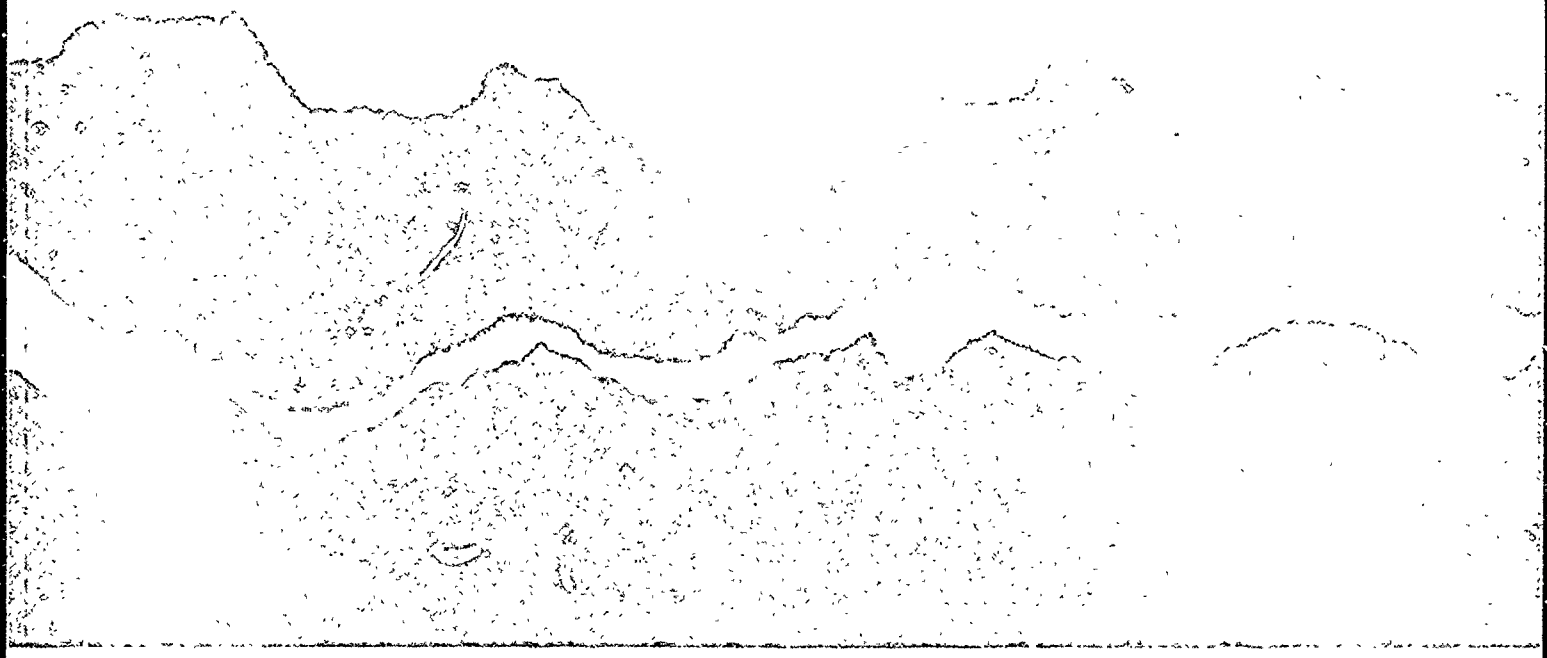
Reprise

In this chapter, a number of issues almost certain to be encountered by evaluators of smoking cessation programs were considered. Because this handbook supplies a number of measures to be used in the evaluation process, special attention was given to the role of such measures in program evaluation. Evaluators desiring more detailed treatments of the topics covered in this chapter will find appropriate sources in the Annotated Bibliography.**

---

* For additional information about reporting the results of an evaluation, see Annotated Bibliography Nos. 5, 23, 26, and 35.

** For additional information about program evaluation, see Annotated Bibliography Nos. 5, 13, 16, 20, 23, 32, 41, 46, 49, and 51.



CHAPTER THREE

Program Evaluation Measures

Overview of Measures

Category	Title	Target Group	Description	Page No.
Behavior	**Smoking Questionnaire**	Adults Adolescents	Assesses use of tobacco products including past smoking history and current level of use.	35
	Smoking History	Adults		37
	Questions About You	Adolescents Preadolescents		40
	Avoiding Smoking	Adults Adolescents	Assesses use of smoking avoidance activities.	44
Knowledge*	**The Physical Effects of Smoking**	Adults Adolescents	Assesses knowledge of effects of smoking on the body.	48
	Facts About Smoking	Adolescents Preadolescents		54
	Smoking and Society	Adults Adolescents	Assesses knowledge of effects of smoking on society.	60
	Problems with Smoking	Adolescents Preadolescents		70
Affective	**Refraining from Smoking**	Adults	Assesses perceived ability to refrain from smoking.	78
	Smoking Situations	Adolescents Preadolescents		82

* The information eligible for inclusion in the knowledge measures is provided in Appendix A as amplified content descriptors.

Category	Title	Target Group	Description	Page No.
Affective	**Beliefs About Smoking**	Adults	Assesses belief in the value of not smoking.	85
	What You Believe About Smoking	Adolescents Preadolescents		89
	Smoking Survey	Adults Adolescents	Assesses intention not to smoke for a specified period of time.	93
	About Smoking	Adolescents Preadolescents		96
	Ideas About Decisions	Adolescents Preadolescents	Assesses belief in the utility of making decisions carefully.	98

SMOKING QUESTIONNAIRE

This behavior measure examines participants' use of tobacco products each day over the last seven days. The questionnaire asks participants to indicate the average number of cigarettes, cigars, or pipefuls of tobacco smoked during the past seven days. Participants are also asked about their use of smokeless tobacco (chewing tobacco, snuff, and nicotine gum). This measure is appropriate for adults and older adolescents.

PURPOSE

Information regarding participants' current tobacco use may be useful for the following reasons:

- Administration of this measure at the beginning of the program may provide needs assessment information. Results from this measure will indicate the extent of the group's smoking or tobacco use prior to program participation. Program personnel will then be able to tailor their program to the participants' smoking level.
- When the measure is administered prior to and following a program, results will demonstrate changes in participants' use of tobacco products.

PROCEDURES

In most cases, this measure should be administered both at the beginning and at the end of the program, particularly if the program is fairly long and emphasizes quitting tobacco use. If the program is short and emphasizes a gradual reduction of tobacco use, it is possible that there will be less change in participants' tobacco use by the end of the program. For programs of shorter duration, program personnel may wish to use this measure for the first purpose described above.

SCORING AND ANALYSIS

This measure can be scored in two ways:

- **Average Daily Use**

  For each tobacco use question, add the responses from all participants and divide this sum by the total number of participants who responded to each question. The resulting score represents the group's average daily use of a particular type of tobacco.

- **Use of Tobacco Products**

  Count the number of participants who indicated that they have used (amounts over 0) any of the five tobacco products listed on the measure. Divide this sum by the total number of respondents, and multiply by 100 to determine the percentage of participants who use tobacco products. This measure can be used for program follow-ups to determine the percentage of participants who are still using tobacco products.

SMOKING QUESTIONNAIRE

This survey asks about your present use of tobacco products.

Think back over the past 7 days.

1. On average, about how many *cigarettes* did you smoke each day?

   _____

2. On average, about how many *cigars* did you smoke each day?

   _____

3. On average, about how many *pipefuls of tobacco* did you smoke each day?

   _____

4. On average, about how many times did you use *chewing tobacco or snuff* each day?

   _____

5. On average, about how many pieces of *nicotine gum* did you chew each day?

   _____

SMOKING HISTORY

This behavior measure examines participants' use of tobacco products over the past few years and in the past 30 days. The measure also examines participants' attempts to quit smoking and how often these attempts were made. Tobacco products referred to on this measure consist of cigarettes, cigars, and pipe tobacco. This measure is appropriate for adults.

PURPOSE

Information regarding participants' past smoking behavior may be useful for needs assessment when collected at the beginning of a program. Results will indicate the group's smoking history as they enter the program. Program personnel will then be able to tailor their program to the participants' past smoking experience.

PROCEDURES

This measure should be administered at the beginning of the program only.

SCORING AND ANALYSIS

The measure can be scored to determine number of years of tobacco use, type of tobacco use, and number of attempts at quitting smoking. Information can also be obtained about the trend of past tobacco use among participants.

- **Years of Smoking (Question 1)**

  To determine the group's average number of years of regular smoking, sum participants' responses to question 1 and divide this total by the number of participants.

- **Past Use (Questions 2 & 3)**

  To determine the average past use of cigarettes, cigars, or pipe tobacco for participants who used these products, sum the responses from all participants for each question. Divide each total by the number of participants who responded to each question. Do not count items that are marked "0" or left blank.

- **Attempts to Quit (Question 4)**

  To determine the average number of unsuccessful attempts to quit smoking for participants who have attempted to quit, sum participants' responses to the second part of question 4 and divide by the number of participants who answered that question.

- **Past Trends (Questions 2 & 3)**

  To compare the group's average past use of each tobacco product (for example, cigarette use), subtract the smallest average score from the largest score for each pair of responses in questions 2 and 3. Next, divide the difference by the average score from question 2 and multiply by 100 to determine the percentage of change in the use of each tobacco product.

If the group's average past use is less in question 3 ("In the past *30 days*") than in question 2, then this would represent a *decrease* in the use of that tobacco product over time. *Increased* use would be represented by a larger group average in question 3 than in question 2.

> *EXAMPLE FOR THE COMPARISON OF PAST USE OF CIGARETTES (Questions 2a and 3a): Imagine that the group's average number of cigarettes recorded in 2a is 20 and in 3a is 15. Subtract 15 from 20 to get 5. Divide 5 by 20 to get .25 which, when multiplied by 100, suggests a 25%* **decrease** *in the level of cigarette smoking from a few years ago to the past 30 days. If, on the other hand, the average in 3a was 24, you would subtract 20 from 24 to get 4. Dividing 4 by 20 and multiplying by 100 results in a 20%* **increase** *in the level of cigarette smoking from a few years ago to the past 30 days.*

SMOKING HISTORY

This survey asks about your use of tobacco in the past.

1. How many years ago did you *first start* smoking tobacco on a regular basis?

   _____

2. In the past *few years*:

   a. About how many cigarettes did you usually smoke *each day*?
   (1 pack = 20 cigarettes)

   _____

   b. About how many cigars did you usually smoke *each day*?

   _____

   c. About how many pipefuls of tobacco did you usually smoke *each day*?

   _____

3. In the past *30 days*:

   a. About how many cigarettes did you usually smoke *each day?*
   (1 pack = 20 cigarettes)

   _____

   b. About how many cigars did you usually smoke *each day*?

   _____

   c. About how many pipefuls of tobacco did you usually smoke *each day*?

   _____

4. Have you ever tried to quit smoking tobacco and found that you couldn't?

   _____

   If yes, how many times have you tried to quit smoking?

   _____

QUESTIONS ABOUT YOU

This behavior measure examines participants' past and present use of tobacco products during the past 30 days and in the past 7 days. This measure is appropriate for adolescents and preadolescents.

PURPOSE

Information about past and current tobacco use may be useful for the following reasons:

- Administration of this measure at the beginning of the program may provide needs assessment information. For example, results of this measure will indicate the group's smoking or tobacco use levels prior to program participation. Program personnel can then tailor their program to meet participants' needs.
- When the measure is administered prior to and at the end of a program, results will demonstrate changes in the frequency with which participants use tobacco products.

PROCEDURES

This measure should be administered both at the beginning and the end of the program.

SCORING AND ANALYSIS

This measure can be scored in two ways:

- **Past Use (Questions 1, 2, 4, and 5)**

  To determine participants' past use of tobacco products, count the number of times each response option is checked in each question for the total group. For example, in question 2 there are 7 response options. Next, divide the sum for each response option by the total number of participants and multiply by 100 to determine the percentage of participants who checked each response concerning past tobacco use.

- **Current Use (Questions 3 & 6)**

  1. To determine participants' current use of each tobacco product, count the number of times "*If none, check here ( )*" is marked for each question. Subtract this total from the total number of program participants. The remainder represents the number of participants who are currently smoking or using chewing tobacco.

2. To determine the average number of cigarettes smoked, total the number of cigarettes participants report smoking for the group (refer to question 3). Next, divide this total by the number of participants who currently smoke cigarettes. Repeat the procedure to determine the average amount of chewing tobacco or snuff used by the group.

By determining current use at the beginning, end, and follow-up stages of the program, program personnel can assess both the number of participants who have quit tobacco use and the average level of use for those who continue to use tobacco.

QUESTIONS ABOUT YOU

Please answer these questions about smoking. Check one answer for each question. To be sure that no one knows how you answered, do *not* write your name on this paper.

1. Have you ever smoked a cigarette?

   ( ) No, never
   ( ) Yes, but only once
   ( ) Yes, more than once

2. How much have you smoked cigarettes during the past 30 days?

   ( ) Not at all
   ( ) Less than one cigarette each day
   ( ) One to five cigarettes each day
   ( ) About one-half pack each day
   ( ) About one pack each day
   ( ) About one and one-half packs each day
   ( ) Two packs or more each day

3. How many cigarettes have you smoked in the last 7 days? *If none, check here* ( ).

   Number of cigarettes _____

4. Have you ever used chewing tobacco or snuff?

   ( ) No, never
   ( ) Yes, but only once
   ( ) Yes, more than once

5. How many times have you used chewing tobacco or snuff during the past 30 days?

   ( ) Never
   ( ) Once
   ( ) Two or three times
   ( ) Once a week
   ( ) Two to four times a week
   ( ) Almost every day
   ( ) Once a day
   ( ) More than once a day

6. How many times have you used chewing tobacco or snuff in the last 7 days? *If none, check here* ( ).

   Number of times _____

AVOIDING SMOKING

This behavior measure examines the frequency with which participants have used a variety of smoking avoidance activities during the past week. This measure is appropriate for adults and adolescents.

PURPOSE

Information about smoking avoidance activities may be useful for the following reasons:

- Administration of this measure at the beginning of the program may provide needs assessment information. For example, results from this measure may indicate the need to broaden participants' array of smoking avoidance activities or may indicate the need to strengthen participants' belief in the value of smoking avoidance activities.
- When given at the beginning and end of a program, results will demonstrate changes in the frequency with which participants successfully use smoking avoidance activities.

PROCEDURES

This measure should be administered both at the beginning and the end of the program.

SCORING AND ANALYSIS

This measure can be scored in two ways:

- **Frequency of use of avoidance activities**

  Count the number of items that are marked OFTEN or SOMETIMES for all participants. (Ignore any blank or NEVER responses.) Divide this total by the number of program participants to determine the average number of smoking avoidance activities used successfully in the past week.

  EXAMPLE: Imagine that there are 10 program participants. First, count all the times that these individuals marked either OFTEN or SOMETIMES. Let's assume that the total number of times was 55. Then, divide 55 by 10 participants to get an average score of 5.5.

  Scores can range from 0-20 with low numbers indicating that the group of participants uses a few smoking avoidance activities successfully and high numbers indicating the successful use of a variety of activities.

- **Frequency of avoidance activities used OFTEN**

  For all participants, count only the items that are marked OFTEN. Divide this total by the number of times the items were marked OFTEN or SOMETIMES. Multiply this number by 100 to obtain a percentage of successful activities used that were marked OFTEN.

To determine the percentage of activities used SOMETIMES, subtract the OFTEN percentage from 100.

EXAMPLE: For the same 10 individuals used in the example above, count the number of times they marked OFTEN. Let's assume the total was 35. Then, divide by the total number of times the 10 individuals marked either OFTEN or SOMETIMES. This number was already determined to be 55 in the previous example. Divide 35 by 55 to find out what percentage of the activities are used OFTEN. In this case, 35/55 is about 64%. Thus, of the activities used successfully, 64% are used often, and 36% are used sometimes.

Besides observing an overall increase in the number of successful smoking avoidance strategies employed, program evaluators using the pretest/posttest approach would hope to see an increase in the frequency with which participants use smoking avoidance activities.

Note: When dealing with the scoring of smoking avoidance activities, program evaluators should not be overly concerned about group scores that do not extend into the upper end of the range. It seems unlikely that even the most skilled participants would use all the smoking avoidance activities listed. Rather, individual participants may find several activities that work well for them.

AVOIDING SMOKING

Listed below are ways that some people avoid smoking. Put a check to show how frequently in the *past week* you successfully used each of these activities *to avoid smoking*.

	Often	Sometimes	Never
1. Exercising	( )	( )	( )
2. Eating or drinking something	( )	( )	( )
3. Thinking about the effort you've put into quitting	( )	( )	( )
4. Chewing gum	( )	( )	( )
5. Using relaxation/deep breathing	( )	( )	( )
6. Calling a friend	( )	( )	( )
7. Giving yourself a "pep talk" not to smoke	( )	( )	( )
8. Promising yourself a reward for not smoking	( )	( )	( )
9. Leaving a situation that makes you want to smoke	( )	( )	( )
10. Thinking about the negative effects of smoking (e.g., poor health, bad breath)	( )	( )	( )
11. Talking to a supportive ex-smoker	( )	( )	( )
12. Keeping busy (e.g., getting involved in a craft or hobby)	( )	( )	( )
13. Reminding yourself of the benefits of not smoking (e.g., better health, money saved)	( )	( )	( )
14. Reading a book or magazine, or watching television	( )	( )	( )
15. Thinking about something besides smoking	( )	( )	( )

	Often	Sometimes	Never
16. Putting off having a cigarette until the urge passes	( )	( )	( )
17. Avoiding places that make you want to smoke	( )	( )	( )
18. Doodling	( )	( )	( )
19. Avoiding frequent contact with people who smoke	( )	( )	( )
20. Giving yourself a reward for not smoking	( )	( )	( )

THE PHYSICAL EFFECTS OF SMOKING
(FORMS A & B)

This knowledge measure examines what participants know about the physical effects of smoking. This measure is appropriate for adults and adolescents.

PURPOSE

Information regarding participants' knowledge of the physical effects of smoking may be useful for the following reasons:

- Administration of this measure at the beginning of the program may provide needs assessment information. For example, the results may be used to assess what participants know prior to program participation. Decisions about how to allocate instructional time can then be made based on the prior knowledge of participants.
- When the measure is administered prior to and following a program, it is possible to evaluate growth in participants' knowledge.

PROCEDURES

Because the equidifficulty of the forms has not been established, it is best not to give all participants Form A as a pretest and Form B as a posttest. Instead, choose either of the following methods.

- Review Forms A and B and select one. Give all participants the selected form both before and after the program. Alternatively, select 20 items from the two forms and construct a measure most consistent with your program emphasis. Then administer the "new" form both before and after the program.
- Give Form A to half of the incoming participants and Form B to the remaining half. To distribute the forms randomly, order them "ABABAB" and hand them out. Following the program, give each participant the form not previously taken. For example, if a participant was given Form B before the program, then that participant should be given Form A following the program. This approach eliminates the possibility that examinees will be sensitized to the specific facts to be learned from the program.

SCORING AND ANALYSIS

The answer keys for the two forms are provided below:

Item No.	Form A	Form B
1	F	T
2	T	F
3	F	T
4	F	T
5	T	F
6	T	T
7	T	T
8	T	T
9	F	F
10	F	F
11	F	T
12	T	T
13	F	F
14	F	F
15	F	T
16	F	F
17	T	F
18	T	F
19	T	F
20	F	T

The measures should be scored by counting the number of correct answers for ea participant. Items marked "Don't Know" or left blank should be scored as inc ct. Next, total the correct answers for the group and divide by the number of participants in the group. The mean number of correct answers and the standard deviation can be used to summarize participant performance on the measure. Means and standard deviations from before and after the program can be compared to determine changes in participants' knowledge.

THE PHYSICAL EFFECTS OF SMOKING

Form A

This test consists of 20 statements about the effects of smoking. Put a check to show whether you think each statement is TRUE or FALSE. If you don't know whether a statement is true or false, put a check under DON'T KNOW.

True	False	Don't Know	
( )	( )	( )	1. Smoking low-tar and low-nicotine cigarettes reduces the risk of all smoking-related diseases.
( )	( )	( )	2. Carbon monoxide is inhaled when a person smokes.
( )	( )	( )	3. How deeply a smoker inhales is not related to that smoker's chance of developing lung cancer.
( )	( )	( )	4. Most experts agree that the harmful effects of smoking on health are not as great for women as for men.
( )	( )	( )	5. Cigarette smoking increases the risk of developing breathing problems.
( )	( )	( )	6. Cigarette smoke can increase the air pollution of homes and offices.
( )	( )	( )	7. Cigarette smoking increases the health dangers associated with taking birth control pills.
( )	( )	( )	8. Frequent pipe and cigar smokers are more likely than nonsmokers to develop lung cancer.
( )	( )	( )	9. The average life expectancy of a smoker is the same as a nonsmoker.
( )	( )	( )	10. People who smoke filter cigarettes inhale less carbon monoxide than people who smoke nonfilter cigarettes.
( )	( )	( )	11. Almost all people gain weight when they quit smoking.
( )	( )	( )	12. Smokers have an increased risk of developing a lung infection after an operation.

True	False	Don't Know	
( )	( )	( )	13. Smoking during pregnancy does not increase the baby's risk of death.
( )	( )	( )	14. Pipe smokers have a greater risk of developing cancer of the mouth than do cigarette smokers.
( )	( )	( )	15. Smoking causes the heart to beat more slowly.
( )	( )	( )	16. The health risks due to smoking do not change even after a person stops smoking.
( )	( )	( )	17. The more a person smokes, the greater is the chance of developing heart disease.
( )	( )	( )	18. Cigarette smoke in the air can cause eye soreness in nonsmokers.
( )	( )	( )	19. On average, babies born to mothers who smoke during pregnancy are smaller than babies born to nonsmokers.
( )	( )	( )	20. Nicotine does *not* cause dependence similar to other addictive drugs.

THE PHYSICAL EFFECTS OF SMOKING

Form B

This test consists of 20 statements about the effects of smoking. Put a check to show whether you think each statement is TRUE or FALSE. If you don't know whether a statement is true or false, put a check under DON'T KNOW.

True	False	Don't Know	
( )	( )	( )	1. Children of smokers have colds and coughs more often than children of nonsmokers.
( )	( )	( )	2. Nicotine causes blood vessels to increase in size.
( )	( )	( )	3. Severe emphysema is a disease rarely found in nonsmokers.
( )	( )	( )	4. About one in every three deaths from cancer is directly related to cigarette smoking.
( )	( )	( )	5. Pipe and cigar smokers are more likely than cigarette smokers to develop cancer of the mouth.
( )	( )	( )	6. A person who has not smoked for ten years has the same chance of developing lung cancer as a person who has never smoked.
( )	( )	( )	7. Cigarette smoking during pregnancy affects the normal growth of the unborn child.
( )	( )	( )	8. Cigarette smokers are about twice as likely as nonsmokers to die of heart disease.
( )	( )	( )	9. Pipe and cigar smoking does not increase a person's chance of developing lung cancer.
( )	( )	( )	10. The unborn child is protected from the effects of the mother smoking.
( )	( )	( )	11. Children are more likely to smoke if their parents smoke.

True	False	Don't Know	
( )	( )	( )	12. Most smokers have made at least one serious attempt to quit smoking.
( )	( )	( )	13. The number of cigarettes smoked by regular smokers is not related to their risk of disease.
( )	( )	( )	14. Pipe smokers who do not inhale have the same chance of developing lung cancer as pipe smokers who do inhale.
( )	( )	( )	15. Many smokers inhale more deeply when they smoke a low-tar and low-nicotine cigarette.
( )	( )	( )	16. Cigarette smoke in the air is not harmful to nonsmokers who breathe it.
( )	( )	( )	17. Smokers of low-tar and low-nicotine cigarettes have the same risk of death as nonsmokers.
( )	( )	( )	18. Only men who smoke have an increased chance of developing lung cancer.
( )	( )	( )	19. Carbon monoxide increases the amount of oxygen in the blood.
( )	( )	( )	20. Cigarette smoking is more damaging to a person's health when combined with exposure to dangerous materials such as asbestos.

FACTS ABOUT SMOKING
(FORMS A & B)

This knowledge measure examines what participants know about the physical effects of smoking. This measure is appropriate for adolescents and preadolescents.

PURPOSE

Information regarding participants' knowledge of the physical effects of smoking may be useful for the following reasons:

- Administration of this measure at the beginning of the program may provide needs assessment information. For example, the results may be used to assess what participants know prior to program participation. Decisions about how to allocate instructional time can then be made based on the prior knowledge of participants.
- When the measure is administered prior to and following a program, it is possible to evaluate growth in participants' knowledge.

PROCEDURES

Because the equidifficulty of the forms has not been established, it is best not to give all participants Form A as a pretest and Form B as a posttest. Instead, choose either of the following methods.

- Review Forms A and B and select one. Give all participants the selected form both before and after the program. Alternatively, select 15 items from the two forms and construct a measure most consistent with your program emphasis. Then administer the "new" form both before and after the program.
- Give Form A to half of the incoming participants and Form B to the remaining half. To distribute the forms randomly, order them "ABABAB" and hand them out. Following the program, give each participant the form not previously taken. For example, if a participant was given Form B before the program, then that participant should be given Form A following the program. This approach eliminates the possibility that examinees will be sensitized to the specific facts to be learned from the program.

SCORING AND ANALYSIS

The answer keys for the two forms are provided below:

Item No.	Form A	Form B
1	F	T
2	T	T
3	F	T
4	T	F
5	F	T
6	T	F
7	T	F
8	T	T
9	F	F
10	F	T
11	T	F
12	T	T
13	F	F
14	T	F
15	F	T

The measures should be scored by counting the number of correct answers for each participant. Items marked "Don't Know" or left blank should be scored as incorrect. Next, total the correct answers for the group and divide by the number of participants in the group. The mean number of correct answers and the standard deviation can be used to summarize participant performance on the measure. Means and standard deviations from before and after the program can be compared to determine changes in participants' knowledge.

FACTS ABOUT SMOKING

Form A

This test contains 15 sentences about smoking. Put a check to show whether you think each sentence is TRUE or FALSE. If you don't know whether a sentence is true or false, put a check under DON'T KNOW.

True	False	Don't Know	
( )	( )	( )	1. Cigarettes that are low in tar are probably safe to smoke.
( )	( )	( )	2. Children whose parents smoke have colds and coughs more often than other children.
( )	( )	( )	3. Smokers and nonsmokers have about the same chance of developing heart disease.
( )	( )	( )	4. Most experts on the effects of smoking think that people who smoke become addicted to nicotine.
( )	( )	( )	5. Smoking cigarettes increases the amount of oxygen in the blood.
( )	( )	( )	6. Babies born to smokers are usually smaller than babies born to nonsmokers.
( )	( )	( )	7. Smokers are more likely to have trouble breathing than nonsmokers.
( )	( )	( )	8. Smokers are usually sick more often than nonsmokers.
( )	( )	( )	9. Cigarette smoke in the air is safe to breathe.
( )	( )	( )	10. Cigarette smokers tend to have lower blood pressure than nonsmokers.
( )	( )	( )	11. Smoking can stain a person's teeth.
( )	( )	( )	12. Smokers have more problems with their gums than nonsmokers.

True	False	Don't Know	
( )	( )	( )	13. Most smokers are able to quit smoking on their first try.
( )	( )	( )	14. People who have quit smoking for ten years have about the same chance of developing lung cancer as nonsmokers.
( )	( )	( )	15. Almost all smokers gain weight when they quit smoking.

FACTS ABOUT SMOKING

Form B

This test contains 15 sentences about smoking. Put a check to show whether you think each sentence is TRUE or FALSE. If you don't know whether a sentence is true or false, put a check under DON'T KNOW.

True	False	Don't Know	
( )	( )	( )	1. Smoking can cause cancer in many parts of the body.
( )	( )	( )	2. People breathe in carbon monoxide when they smoke.
( )	( )	( )	3. Many people would like to stop smoking but find that they can't.
( )	( )	( )	4. Low-tar cigarettes are probably safe to smoke.
( )	( )	( )	5. A woman who smokes while pregnant increases the chance that her baby will be harmed.
( )	( )	( )	6. Smokers and nonsmokers have about the same chance of having a heart attack.
( )	( )	( )	7. Smoking cigarettes makes it more likely for men, but not women, to develop lung cancer.
( )	( )	( )	8. Smokers have coughs and colds more often than nonsmokers.
( )	( )	( )	9. Smoking cigarettes makes the heart beat slower.
( )	( )	( )	10. Nicotine can produce a drug-like dependence in smokers.
( )	( )	( )	11. Smokers usually live as long as nonsmokers.
( )	( )	( )	12. Smoking can give you bad breath.
( )	( )	( )	13. Smoking just a few cigarettes every day is safe.

True	False	Don't Know	
( )	( )	( )	14. Cigarettes that have filters are safe to smoke.
( )	( )	( )	15. Cigarette smoke in the air may be bad for people who breathe it.

SMOKING AND SOCIETY
(FORMS A & B)

This knowledge measure examines what participants know about the effects of smoking on society. This measure is appropriate for adults and adolescents.

PURPOSE

Information regarding participants' knowledge of the effects of smoking on society may be useful for the following reasons:

- Administration of this measure at the beginning of the program may provide needs assessment information. For example, the results may be used to assess what participants know prior to program participation. Decisions about how to allocate instructional time can then be made based on the prior knowledge of participants.
- When the measure is administered prior to and following a program, it is possible to evaluate growth in participants' knowledge.

PROCEDURES

Because the equidifficulty of the forms has not been established, it is best not to give all participants Form A as a pretest and Form B as a posttest. Instead, choose either of the following methods.

- Review Forms A and B and select one. Give all participants the selected form both before and after the program. Alternatively, select 15 items from the two forms and construct a measure most consistent with your program emphasis. Then administer the "new" form both before and after the program.
- Give Form A to half of the incoming participants and Form B to the remaining half. To distribute the forms randomly, order them "ABABAB" and hand them out. Following the program, give each participant the form not previously taken. For example, if a participant was given Form B before the program, then that participant should be given Form A following the program. This approach eliminates the possibility that examinees will be sensitized to the specific facts to be learned from the program.

SCORING AND ANALYSIS

The answer keys for the two forms are provided below:

Item No.	Form A	Form B
1	C	B
2	A	A
3	B	B
4	A	C
5	C	B
6	B	A
7	C	C
8	B	B
9	A	C
10	C	C
11	B	A
12	A	B
13	C	C
14	B	A
15	A	C

The measures should be scored by counting the number of correct answers for each participant. Items marked "Don't Know" or left blank should be scored as incorrect. Next, total the correct answers for the group and divide by the number of participants in the group. The mean number of correct answers and the standard deviation can be used to summarize participant performance on the measure. Means and standard deviations from before and after the program can be compared to determine changes in participants' knowledge.

SMOKING AND SOCIETY

Form A

This test consists of 15 questions about smoking in American society. Circle one answer for each question. If you are unsure of what the correct answer is, circle D for DON'T KNOW.

1. How much are the health consequences of smoking estimated to cost in the United States each year?

   A. $500 million

   B. $15 billion

   C. $30 billion

   D. Don't know

2. Which of the following is true about the difficulty of quitting smoking for men and women?

   A. Women find it more difficult to quit than men.

   B. Men find it more difficult to quit than women.

   C. Men and women have about the same difficulty quitting.

   D. Don't know

3. About how many people annually are estimated to die prematurely from smoking?

   A. 30 thousand

   B. 300 thousand

   C. 3 million

   D. Don't know

4. Which of the following is true about the effects of environmental smoke on the children of parents who smoke?

   A. They have an increased risk of hospitalization for bronchitis and pneumonia.

   B. They have increased difficulty with normal eating and sleeping patterns.

   C. They have no detectable difficulties that set them apart from other children.

   D. Don't know

5. Which of the following is true about the effects of involuntary or environmental smoke on nonsmokers?

   A. Nonsmokers with allergies to smoke are the only nonsmokers with demonstrated negative effects.

   B. There are no demonstrated negative effects on nonsmokers.

   C. Environmental tobacco smoke can cause lung cancer in healthy nonsmokers.

   D. Don't know

6. Which of the following is true about the filtration of tobacco smoke from the air?

   A. Very simple, cost-effective methods exist for filtering tobacco smoke particles from the air.

   B. Effective removal of smoke particles from indoor air requires an increase in the exchange with outdoor air.

   C. No method currently exists for lowering the number of tobacco smoke particles in indoor air.

   D. Don't know

7. Of the approximately 60,000 people who die each year from chronic obstructive lung disease, what percentage can be attributed to smoking?

   A. 30% to 40%

   B. 50% to 60%

   C. 80% to 90%

   D. Don't know

8. Which one of the following smoking-related diseases is estimated to account for the most deaths in the American population?

   A. Smoking-related lung cancer

   B. Smoking-related cardiovascular disease

   C. Smoking-related emphysema

   D. Don't know

9. Which of the following is true concerning the health risks to the fetuses of mothers who smoke?

   A. Maternal smoking contributes to prenatal mortality and low birth weight.

   B. Because smoke is filtered by the mother's body, there are few health risks to the fetus.

   C. There is little conclusive evidence that there are any health risks to the fetus.

   D. Don't know

10. Which of the following is true about the costs of smoking in comparison to the costs of drug and alcohol abuse?

    A. The effects of smoking cost society less than the costs of alcohol or drug abuse.

    B. The effects of smoking cost society more than drug abuse but less than alcohol abuse.

    C. The effects of smoking cost society more than the cost of alcohol abuse or drug abuse.

    D. Don't know

11. Of the more than 135,000 lung cancer deaths per year in the United States, what percentage are directly attributable to cigarette smoking?

    A. 60%

    B. 85%

    C. 95%

    D. Don't know

12. Approximately what percentage of the current American adult population smokes?
    A. 25%
    B. 50%
    C. 75%
    D. Don't know

13. In families where both parents smoke, about what percentage of the adolescent children also smoke?
    A. 5%
    B. 15%
    C. 25%
    D. Don't know

14. Which of the following is true about patterns in smoking of teenage boys and girls?
    A. The number of both teenage boys and teenage girls who smoke is rising rapidly each year.
    B. The number of teenage boys who smoke is staying about the same while the number of girls is going up.
    C. The number of teenage boys and girls who smoke is gradually going down each year.
    D. Don't know

15. Which of the following is the most likely cause of adolescents beginning smoking?
    A. Social influences
    B. Stress
    C. Ignorance of smoking effects
    D. Don't know

SMOKING AND SOCIETY

Form B

This test consists of 15 questions about smoking in American society. Circle one answer for each question. If you are unsure of what the correct answer is, circle D for DON'T KNOW.

1. Which of the following is true about the smoking patterns of men and women?

   A. More women smoke than men.

   B. More men smoke than women.

   C. About the same number of men and women smoke.

   D. Don't know

2. Which of the following is true about the effect of separating smokers and nonsmokers within the same air space?

   A. It may reduce but does not eliminate nonsmokers' exposure to environmental tobacco smoke.

   B. It eliminates all significant exposure to environmental tobacco smoke.

   C. It does nothing to reduce the amount of exposure that nonsmokers receive from environmental tobacco smoke.

   D. Don't know

3. How much higher is the death rate from coronary heart disease for smokers than for nonsmokers?

   A. 50%

   B. 70%

   C. 90%

   D. Don't know

4. Which of the following is *not* true about coronary heart disease?

   A. Coronary heart disease is the single most important cause of death in the United States.

   B. Cigarette smoking ranks as the largest preventable cause of coronary heart disease.

   C. When all factors are weighed together, smoking is the only preventable cause of coronary heart disease.

   D. Don't know

5. Approximately what percentage of all deaths in the United States are related to smoking?

   A. 5%

   B. 15%

   C. 25%

   D. Don't know

6. Which of the following is true about the difference between sidestream smoke and mainstream smoke from a cigarette?

   A. Greater amounts of some carcinogens are found in sidestream smoke than in mainstream smoke.

   B. About the same amounts of carcinogens are found in sidestream smoke as in mainstream smoke.

   C. Greater amounts of carcinogens are found in mainstream smoke than in sidestream smoke.

   D. Don't know

7. For every dollar Americans spend purchasing cigarettes, about how much money is spent directly on health care costs for smoking-related diseases?

   A. 25 cents

   B. 50 cents

   C. 1 dollar

   D. Don't know

8. What has been the effect of the published reports since the 1960s on the health risks of smoking?

   A. The percentage of smokers quitting has decreased.

   B. The percentage of smokers quitting has increased.

   C. The percentage of smokers quitting has stayed about the same.

   D. Don't know

9. Which of the following is true about the effects of mothers quitting smoking during pregnancy?

   A. The damaging effects of smoking on fetal development are not reduced by the mother quitting smoking during pregnancy.

   B. The damaging effects of smoking on fetal development are reduced only if the mother stops smoking at least 6 months before becoming pregnant.

   C. The damaging effects of smoking on fetal development are reduced if the mother stops smoking during her pregnancy.

   D. Don't know

10. Which of the following is true about smokers' risk of death from lung cancer?

    A. Smokers' risk of death from lung cancer is roughly the same as that of nonsmokers.

    B. Smokers' risk of death from lung cancer is roughly three times greater than that of nonsmokers.

    C. Smokers' risk of death from lung cancer is roughly ten times greater than that of nonsmokers.

    D. Don't know

11. About what percentage of adolescents ages 12 - 14 smoke once a week or more?

    A. 5%

    B. 15%

    C. 25%

    D. Don't know

12. Cigarettes start fires that account for what percentage of all fire-related deaths in the United States?

    A. 10%

    B. 30%

    C. 50%

    D. Don't know

13. Approximately how much do Americans spend each year purchasing cigarettes?

    A. $250 million

    B. $1 billion

    C. $30 billion

    D. Don't know

14. Which of the following is true about the smoking behavior of children in families where both parents smoke?

    A. Twice as many children smoke as in families where neither parent smokes.

    B. About the same number smoke as in families where neither parent smokes.

    C. Fewer children smoke than in families where neither parent smokes.

    D. Don't know

15. Which of the following is true about the gap between the number of women and men who smoke?

    A. The gap is becoming wider.

    B. The gap is staying about the same.

    C. The gap is becoming narrower.

    D. Don't know

PROBLEMS WITH SMOKING
(FORMS A & B)

This knowledge measure examines what participants know about the effects of smoking on society. This measure is appropriate for adolescents and preadolescents.

PURPOSE

Information regarding participants' knowledge of the effects of smoking on society may be useful for the following reasons:

- Administration of this measure at the beginning of the program may provide needs assessment information. For example, the results may be used to assess what participants know prior to program participation. Decisions about how to allocate instructional time can then be made based on the prior knowledge of participants.
- When the measure is administered prior to and following a program, it is possible to evaluate growth in participants' knowledge.

PROCEDURES

Because the equidifficulty of the forms has not been established, it is best not to give all participants Form A as a pretest and Form B as a posttest. Instead, choose either of the following methods.

- Review Forms A and B and select one. Give all participants the selected form both before and after the program. Alternatively, select 10 items from the two forms and construct a measure most consistent with your program emphasis. Then administer the "new" form both before and after the program.
- Give Form A to half of the incoming participants and Form B to the remaining half. To distribute the forms randomly, order them "ABABAB" and hand them out. Following the program, give each participant the form not previously taken. For example, if a participant was given Form B before the program, then that participant should be given Form A following the program. This approach eliminates the possibility that examinees will be sensitized to the specific facts to be learned from the program.

SCORING AND ANALYSIS

The answer keys for the two forms are provided below:

Item No.	Form A	Form B
1	C	A
2	C	A
3	B	B
4	A	B
5	B	A
6	C	C
7	B	C
8	A	C
9	B	B
10	A	C

The measures should be scored by counting the number of correct answers for each participant. Items marked "Don't Know" or left blank should be scored as incorrect. Next, total the correct answers for the group and divide by the number of participants in the group. The mean number of correct answers and the standard deviation can be used to summarize participant performance on the measure. Means and standard deviations from before and after the program can be compared to determine changes in participants' knowledge.

PROBLEMS WITH SMOKING

Form A

This test contains 10 questions about smoking in America. Circle one answer for each question. If you are unsure of what the correct answer is, circle D for DON'T KNOW.

1. About how much do the health problems caused by smoking cost in the United States each year?

   A. $500 million

   B. $15 billion

   C. $30 billion

   D. Don't know

2. In families where both parents smoke, about what percentage of the children also smoke?

   A. 5%

   B. 15%

   C. 25%

   D. Don't know

3. As a result of smoking, how many people die each year before they normally would?

   A. 30 thousand

   B. 300 thousand

   C. 3 million

   D. Don't know

4. Which of the following is true about the result of breathing tobacco smoke in the air?

   A. Tobacco smoke in the air can cause lung cancer in healthy nonsmokers.

   B. Nonsmokers who are allergic to smoke are the only nonsmokers who have health problems caused by tobacco smoke in the air.

   C. Nonsmokers have no serious health problems from tobacco smoke in the air.

   D. Don't know

5. Which of the following is true about the number of teenage boys and girls who smoke?

   A. The number of both teenage boys and teenage girls who smoke is rising very rapidly each year.

   B. The number of teenage boys who smoke is staying about the same while the number of girls who smoke is rising.

   C. The number of teenage boys and girls who smoke is gradually going down each year.

   D. Don't know

6. Of the roughly 60,000 people who die each year from emphysema and other related lung diseases, what percentage of these deaths are caused by smoking?

   A. 30% to 40%

   B. 50% to 60%

   C. 80% to 90%

   D. Don't know

7. Which one of the following smoking-related diseases is estimated to account for the most deaths in the America?

   A. Smoking-related lung cancer

   B. Smoking-related heart disease

   C. Smoking-related emphysema

   D. Don't know

8. About what percentage of children ages 12-14 smoke once a week or more?

   A. 5%

   B. 15%

   C. 25%

   D. Don't know

9. Of the more than 135,000 deaths from lung cancer per year in the United States, what percentage are directly related to cigarette smoking?

   A. 60%

   B. 85%

   C. 95%

   D. Don't know

10. About what percentage of American adults smoke?

    A. 25%

    B. 50%

    C. 75%

    D. Don't know

PROBLEMS WITH SMOKING

Form B

This test contains 10 questions about smoking in America. Circle one answer for each question. If you are unsure of what the correct answer is, circle D for DON'T KNOW.

1. Which of the following is true about the smoking behavior of children in families where both parents smoke?

   A. Twice as many children smoke as in families where neither parent smokes.

   B. About the same number smoke when compared to families where neither parent smokes.

   C. Fewer children smoke than in families where neither parent smokes.

   D. Don't know

2. Which of the following is true when nonsmokers are kept away from people who are smoking within the same area?

   A. It reduces but does not do away with the smoke that nonsmokers breathe.

   B. Nonsmokers are not exposed to any of the harmful elements in cigarette smoke.

   C. It does nothing to keep nonsmokers from being exposed to harmful cigarette smoke.

   D. Don't know

3. How much higher is the death rate from heart disease for smokers than for nonsmokers?

   A. 50% higher

   B. 70% higher

   C. 90% higher

   D. Don't know

4. About what percentage of all deaths in the United States are related to smoking?

   A. 5%

   B. 15%

   C. 25%

   D. Don't know

5. Which of the following is true about the difference between the smoke from the burning end of a cigarette (sidestream smoke) and smoke that the smoker breathes in through the cigarette (mainstream smoke)?

   A. Greater amounts of some cancer-causing elements are found in sidestream smoke than in mainstream smoke.

   B. About the same amounts of cancer-causing elements are found in sidestream smoke as in mainstream smoke.

   C. Greater amounts of cancer-causing elements are found in mainstream smoke than in sidestream smoke.

   D. Don't know

6. Which of the following is true about children who experiment with smoking?

   A. Children who experiment with smoking usually do not grow up to be regular smokers as adults.

   B. Children who experiment with smoking have about the same chance of becoming smokers as those who do not experiment.

   C. Children who experiment with smoking often grow up to be regular smokers as adults.

   D. Don't know

7. Which of the following is true about what can happen to an unborn baby if its mother smokes?

   A. A mother's smoking usually causes no problems for the unborn baby.

   B. A mother's smoking often causes the baby to weigh more when it is born.

   C. A mother's smoking causes a greater chance of the baby dying before birth.

   D. Don't know

8. Which of the following is true about a smoker's risk of death from lung cancer?

   A. A smoker's risk of death from lung cancer is about the same as that of a nonsmoker.

   B. A smoker's risk of death from lung cancer is about three times greater than that of a nonsmoker.

   C. A smoker's risk of death from lung cancer is about ten times greater than that of a nonsmoker.

   D. Don't know

9. Cigarettes start fires that are responsible for what percentage of all fire-related deaths in the United States?

   A. 10%

   B. 30%

   C. 50%

   D. Don't know

10. About how much do Americans spend each year buying cigarettes?

   A. $250 million

   B. $1 billion

   C. $30 billion

   D. Don't know

REFRAINING FROM SMOKING

This affective measure assesses participants' perceptions regarding their ability to refrain from smoking. This measure is appropriate for adults.

PURPOSE

Having affective information about participants' perceptions regarding their ability to refrain from smoking may be useful for the following reasons:

- Administration of this measure at the beginning of the program may provide needs assessment information. For example, results of this measure may indicate a lack of perceived ability to refrain from smoking and thus indicate a need for participant training in that area.
- When this measure is administered prior to and following a program it is possible to evaluate changes in participants' perceptions regarding their ability to refrain from smoking.

PROCEDURES

This instrument can be administered both at the beginning and at the end of the program. However, handbook users should be alert to concerns regarding the potential reactivity of affective measures. A measure is considered *reactive* if the experience of completing the measure prior to the program causes participants to react differently to the program. Handbook users should, therefore, carefully review each affective measure that they wish to use to determine its potential for making participants unduly sensitive to aspects of the program. If a measure is determined to be reactive, then program personnel should *not* administer that measure to *all* participants as a pretest and posttest. Instead, the measure could be administered to half of the program participants prior to program participation to determine participants' pre-program status. The measure could then be administered to the other half of the participants after program participation to assess participants' post-program status.

SCORING AND ANALYSIS

Point values are assigned to responses as follows:

Definitely Yes	=	5
Probably Yes	=	4
Maybe	=	3
Probably No	=	2
Definitely No	=	1

This inventory can be scored by adding the point values of the responses from all participants and dividing this total by the number of responses. Blank items should not be counted in the number of responses. The maximum attainable score of 5 points indicates a strong perceived ability to refrain from smoking across a variety of potential smoking situations. A minimum score of 1 indicates a perceived lack of ability to refrain from smoking in a variety of situations.

REFRAINING FROM SMOKING

This survey describes times when people often feel an urge to smoke. Put a check to show how sure you are that you could refrain or keep from smoking in each situation.

Could you refrain from smoking if ...	Definitely Yes	Probably Yes	Maybe	Probably No	Definitely No
1. you had just finished an enjoyable meal?	( )	( )	( )	( )	( )
2. you were drinking coffee or tea?	( )	( )	( )	( )	( )
3. you were watching television?	( )	( )	( )	( )	( )
4. you were visiting friends, some of whom were smoking?	( )	( )	( )	( )	( )
5. you had just completed a difficult task that had taken you a long time to finish?	( )	( )	( )	( )	( )
6. you were tense and anxious?	( )	( )	( )	( )	( )
7. you were reading a newspaper or magazine?	( )	( )	( )	( )	( )
8. you were talking on the telephone?	( )	( )	( )	( )	( )
9. you just had a big argument with someone in your family?	( )	( )	( )	( )	( )
10. you were relaxing after a busy day?	( )	( )	( )	( )	( )

Could you refrain from smoking if ...	Definitely Yes	Probably Yes	Maybe	Probably No	Definitely No
11. you hadn't had a cigarette in a while and someone offered you one?	( )	( )	( )	( )	( )
12. you were waiting for a very important phone call that was fifteen minutes late?	( )	( )	( )	( )	( )
13. you wanted to avoid eating sweets?	( )	( )	( )	( )	( )
14. you were at a party and someone offered you a cigarette?	( )	( )	( )	( )	( )
15. you were at a sporting or entertainment event?	( )	( )	( )	( )	( )
16. you felt as if you really needed to smoke?	( )	( )	( )	( )	( )
17. you were taking a work break?	( )	( )	( )	( )	( )
18. you were with a friend who urged you to smoke?	( )	( )	( )	( )	( )
19. you were tired and needed more energy?	( )	( )	( )	( )	( )
20. you were driving to work in the morning?	( )	( )	( )	( )	( )
21. you were having a few drinks with friends in a bar or cocktail lounge?	( )	( )	( )	( )	( )
22. you were alone and feeling depressed?	( )	( )	( )	( )	( )
23. you were celebrating a special occasion?	( )	( )	( )	( )	( )

Could you refrain from smoking if ...	Definitely Yes	Probably Yes	Maybe	Probably No	Definitely No
24. you were doing paperwork such as studying, paying bills, or writing a letter?	( )	( )	( )	( )	( )
25. you noticed that you were starting to put on weight?	( )	( )	( )	( )	( )
26. you wanted to feel more sophisticated and attractive?	( )	( )	( )	( )	( )
27. you were bored?	( )	( )	( )	( )	( )
28. Could you refrain from smoking regardless of the circumstances?	( )	( )	( )	( )	( )

SMOKING SITUATIONS

This affective measure assesses participants' perceptions regarding their ability to refrain from smoking. This measure is appropriate for adolescents and preadolescents.

PURPOSE

Having affective information about participants' perceptions regarding their ability to refrain from smoking may be useful for the following reasons:

- Administration of this measure at the beginning of the program may provide needs assessment information. For example, results of this measure may indicate a lack of perceived ability to refrain from smoking and thus indicate a need for participant training in that area.
- When this measure is administered prior to and following a program, it is possible to evaluate changes in participants' perceptions regarding their ability to refrain from smoking.

PROCEDURES

This instrument can be administered both at the beginning and at the end of the program. However, handbook users should be alert to concerns regarding the potential reactivity of affective measures. A measure is considered *reactive* if the experience of completing the measure prior to the program causes participants to react differently to the program. Handbook users should, therefore, carefully review each affective measure that they wish to use to determine its potential for making participants unduly sensitive to aspects of the program. If a measure is determined to be reactive, then program personnel should *not* administer that measure to *all* participants as a pretest and posttest. Instead, the measure could be administered to half of the program participants prior to program participation to determine participants' pre-program status. The measure could then be administered to the other half of the participants after program participation to assess participants' post-program status.

SCORING AND ANALYSIS

Point values are assigned to responses as follows:

Definitely Yes	=	5
Probably Yes	=	4
Maybe	=	3
Probably No	=	2
Definitely No	=	1

This inventory can be scored by adding the point values of the responses from all participants and dividing this total by the number of responses. Blank items should not be counted in the number of responses. The maximum attainable score of 5 points indicates a strong perceived ability to refrain from smoking across a variety of potential smoking situations. A minimum score of 1 indicates a perceived lack of ability to refrain from smoking in a variety of situations.

SMOKING SITUATIONS

Young people sometimes find themselves in situations in which they feel pressure to smoke. Some of these situations are described below. Put a check to show how sure you are that you could *keep from smoking* in each situation.

	Definitely Yes	Probably Yes	Maybe	Probably No	Definitely No
1. You're invited to a party with the most popular kids at school. Many people are smoking. As you talk in a small group, someone offers you a puff. Could you keep from smoking?	( )	( )	( )	( )	( )
2. You go to a friend's house to study. He suggests that you both try a cigarette. No one but your friend would know. Could you keep from smoking?	( )	( )	( )	( )	( )
3. You're at a football game with a new friend. Her friends are passing around a cigarette. Your friend takes a puff and hands it to you. Could you keep from smoking?	( )	( )	( )	( )	( )
4. Your older sister hides cigarettes in her room. You're all alone at home. It would be easy to try one. Could you keep from smoking?	( )	( )	( )	( )	( )
5. You're watching T.V. at your uncle's house. He joins you and lights up a cigarette. He's in a good mood and jokingly offers you a puff. You know he'll tease you if you don't give it a try. Could you keep from smoking?	( )	( )	( )	( )	( )

	Definitely Yes	Probably Yes	Maybe	Probably No	Definitely No
6. You're at a dance and have met someone you think is really nice. When you take a walk outside you find out your new friend smokes. Could you keep from smoking?	( )	( )	( )	( )	( )
7. You're walking home from school with some friends. One of them passes a pack of cigarettes around and everybody takes one. Could you keep from smoking?	( )	( )	( )	( )	( )
8. You decide to have a party on a weekend that your parents are gone. Your best friend brings some cigarettes to have around in case people want to smoke. Later, it seems like a lot of people are smoking. Could you keep from smoking?	( )	( )	( )	( )	( )
9. During lunch your friends go to the edge of the school grounds to smoke together. You don't want to be left out of the group. Could you keep from smoking?	( )	( )	( )	( )	( )
10. You've just moved to a new neighborhood. A group of kids at your new school have been really nice to you. You would like to be part of their group. Most of them smoke. Could you keep from smoking?	( )	( )	( )	( )	( )

BELIEFS ABOUT SMOKING

This affective measure assesses participants' belief in the value of not smoking. This measure is appropriate for adults.

PURPOSE

Information about participants smoking beliefs may be useful for the following reasons:

- Administration of this measure at the beginning of the program may provide needs assessment information. For example, results of this measure may indicate a need for assisting participants in strengthening their beliefs about the negative effects of smoking.
- When this measure is administered prior to and following a program, it is possible to evaluate changes in participants' beliefs about the negative effects of smoking.

PROCEDURES

This instrument can be administered both at the beginning and at the end of the program. However, handbook users should be alert to concerns regarding the potential reactivity of affective measures. A measure is considered *reactive* if the experience of completing the measure prior to the program causes participants to react differently to the program. Handbook users should, therefore, carefully review each affective measure that they wish to use to determine its potential for making participants unduly sensitive to aspects of the program. If a measure is determined to be reactive, then program personnel should *not* administer that measure to *all* participants as a pretest and posttest. Instead, the measure could be administered to half of the program participants prior to program participation to determine participants' pre-program status. The measure could then be administered to the other half of the participants after program participation to assess participants' post-program status.

SCORING AND ANALYSIS

Point values are assigned to responses according to the following scoring key:

Item No.	Strongly Agree	Agree	Not Sure	Disagree	Strongly Disagree
1	1	2	3	4	5
2	5	4	3	2	1
3	1	2	3	4	5
4	1	2	3	4	5
5	1	2	3	4	5
6	1	2	3	4	5
7	1	2	3	4	5
8	1	2	3	4	5
9	1	2	3	4	5
10	1	2	3	4	5
11	1	2	3	4	5
12	1	2	3	4	5
13	5	4	3	2	1
14	5	4	3	2	1
15	1	2	3	4	5

This inventory can be scored by adding the point values of the responses from all participants and dividing this total by the number of responses. Blank items should not be counted in the number of responses. The maximum attainable score of 5 points indicates a strong belief in the negative effects of smoking. A minimum score of 1 suggests weak belief in the negative effects of smoking.

BELIEFS ABOUT SMOKING

The sentences below are about how you might be affected by smoking. Put a check to show how much you agree or disagree with each sentence.

	Strongly Agree	Agree	Not Sure	Disagree	Strongly Disagree
1. I would have to smoke regularly for many years before smoking would affect my health.	( )	( )	( )	( )	( )
2. Smoking only a few cigarettes a day would hurt my health.	( )	( )	( )	( )	( )
3. People who smoke are more successful than those who don't smoke.	( )	( )	( )	( )	( )
4. Pipes and cigars are safe to smoke if I don't inhale.	( )	( )	( )	( )	( )
5. After quitting, ex-smokers will be as healthy as if they had never smoked.	( )	( )	( )	( )	( )
6. The health risks of smoking can be overcome through exercise.	( )	( )	( )	( )	( )
7. Social gatherings are better when people are smoking.	( )	( )	( )	( )	( )
8. Smoking cigarettes is a sign of being mature.	( )	( )	( )	( )	( )
9. Most people who smoke cigarettes can quit smoking whenever they want to.	( )	( )	( )	( )	( )

	Strongly Agree	Agree	Not Sure	Disagree	Strongly Disagree
10. The health risks of smoking have been greatly exaggerated.	( )	( )	( )	( )	( )
11. People enjoy life more when they smoke.	( )	( )	( )	( )	( )
12. Weight gain is an unavoidable result of quitting smoking.	( )	( )	( )	( )	( )
13. Most people who smoke cigarettes want to quit.	( )	( )	( )	( )	( )
14. Laws that limit advertising for cigarettes should be enforced.	( )	( )	( )	( )	( )
15. Smoking helps me through stressful situations.	( )	( )	( )	( )	( )

WHAT YOU BELIEVE ABOUT SMOKING

This affective measure assesses participants' belief in the value of not smoking. This measure is appropriate for adolescents and preadolescents.

PURPOSE

Information about participants' smoking beliefs may be useful for the following reasons:

- Administration of this measure at the beginning of the program may provide needs assessment information. For example, results of this measure may indicate a need for assisting participants in strengthening their beliefs about the negative effects of smoking.
- When this measure is administered prior to and following a program, it is possible to evaluate changes in participants' beliefs about the negative effects of smoking.

PROCEDURES

This instrument can be administered both at the beginning and at the end of the program. However, handbook users should be alert to concerns regarding the potential reactivity of affective measures. A measure is considered *reactive* if the experience of completing the measure prior to the program causes participants to react differently to the program. Handbook users should, therefore, carefully review each affective measure that they wish to use to determine its potential for making participants unduly sensitive to aspects of the program. If a measure is determined to be reactive, then program personnel should *not* administer that measure to *all* participants as a pretest and posttest. Instead, the measure could be administered to half of the program participants prior to program participation to determine participants' pre-program status. The measure could then be administered to the other half of the participants after program participation to assess participants' post-program status.

SCORING AND ANALYSIS

Point values are assigned to responses according to the following scoring key:

Item No.	Strongly Agree	Agree	Not Sure	Disagree	Strongly Disagree
1	1	2	3	4	5
2	5	4	3	2	1
3	1	2	3	4	5
4	1	2	3	4	5
5	1	2	3	4	5
6	1	2	3	4	5
7	1	2	3	4	5
8	1	2	3	4	5
9	1	2	3	4	5
10	1	2	3	4	5
11	1	2	3	4	5
12	1	2	3	4	5
13	5	4	3	2	1
14	5	4	3	2	1
15	1	2	3	4	5

This inventory can be scored by adding the point values of the responses from all participants and dividing this total by the number of responses. Blank items should not be counted in the number of responses. The maximum attainable score of 5 points indicates a strong belief in the negative effects of smoking. A minimum score of 1 suggests weak belief in the negative effects of smoking.

WHAT YOU BELIEVE ABOUT SMOKING

The sentences below are about smoking. Put a check to show how much you agree or disagree with each sentence.

	Strongly Agree	Agree	Not Sure	Disagree	Strongly Disagree
1. I would have to smoke for a long time before it would hurt my health.	( )	( )	( )	( )	( )
2. Smoking only a few cigarettes a day would hurt my health.	( )	( )	( )	( )	( )
3. People who smoke are more popular than those who don't smoke.	( )	( )	( )	( )	( )
4. It would be safe to smoke cigarettes if I didn't inhale.	( )	( )	( )	( )	( )
5. After quitting, an ex-smoker's health will be as good as it ever was.	( )	( )	( )	( )	( )
6. Smoking would not hurt my health if I exercised a lot.	( )	( )	( )	( )	( )
7. Parties are better when people are smoking.	( )	( )	( )	( )	( )
8. Smoking cigarettes is part of growing up.	( )	( )	( )	( )	( )
9. Most teenagers who smoke cigarettes can stop smoking whenever they want to.	( )	( )	( )	( )	( )
10. Smoking is not as bad for your health as some people make it seem.	( )	( )	( )	( )	( )
11. People have more fun when they smoke.	( )	( )	( )	( )	( )

	Strongly Agree	Agree	Not Sure	Disagree	Strongly Disagree
12. I could smoke cigarettes without getting hooked.	( )	( )	( )	( )	( )
13. Most adults who smoke cigarettes want to quit.	( )	( )	( )	( )	( )
14. There should be laws that limit advertising for cigarettes.	( )	( )	( )	( )	( )
15. People who smoke are better athletes than those who don't smoke.	( )	( )	( )	( )	( )

SMOKING SURVEY

This affective measure assesses participants' intention to refrain from smoking or to quit smoking. This measure is appropriate for adults and adolescents.

PURPOSE

Information about participants' intention to refrain from smoking or to quit smoking may be useful for the following reasons:

- Administration of this measure at the beginning of the program may provide needs assessment information. For example, results of this measure may indicate that participants have little intention to refrain from smoking, thus emphasizing a need for instruction in that area.
- When this measure is administered prior to and following a program, the results will demonstrate the program's effects on participants' intention to refrain from smoking.

PROCEDURES

This instrument can be administered both at the beginning and at the end of the program. However, handbook users should be alert to concerns regarding the potential reactivity of affective measures. A measure is considered *reactive* if the experience of completing the measure prior to the program causes participants to react differently to the program. Handbook users should, therefore, carefully review each affective measure that they wish to use to determine its potential for making participants unduly sensitive to aspects of the program. If a measure is determined to be reactive, then program personnel should *not* administer that measure to *all* participants as a pretest and posttest. Instead, the measure could be administered to half of the program participants prior to program participation to determine participants' pre-program status. The measure could then be administered to the other half of the participants after program participation to assess participants' post-program status.

SCORING AND ANALYSIS

The measure consists of two sets of questions, one for current smokers and one for current nonsmokers. The two columns should be scored separately. Point values are assigned to responses as follows:

Definitely Yes (A)	=	5
Probably Yes (B)	=	4
Maybe (C)	=	3
Probably No (D)	=	2
Definitely No (E)	=	1

For either column, add the point values of the responses from all participants and divide this total by the number of responses. Blank items should not be counted in the number of responses. The maximum attainable score of 5 points indicates a strong intention to refrain from (or quit) smoking. A minimum score of 1 suggests a weak intention to refrain from (or quit) smoking.

SMOKING SURVEY

This survey asks about your plans to quit or refrain from (avoid) smoking. First, indicate whether or not you currently smoke. Then, use the following scale to answer the questions under the box you check.

A	B	C	D	E
Definitely Yes	Probably Yes	Maybe	Probably No	Definitely No

CHECK ONE BOX:

Do you currently smoke tobacco?

☐ **No**

1. Do you plan to refrain from smoking throughout the next week? (Circle one)

   A B C D E

2. Do you plan to refrain from smoking throughout the next month? (Circle one)

   A B C D E

3. Do you plan to refrain from smoking throughout the next year? (Circle one)

   A B C D E

4. Do you plan to refrain from smoking for the rest of your life? (Circle one)

   A B C D E

☐ **Yes**

1. Do you plan to quit smoking within the next week? (Circle one)

   A B C D E

2. Are you likely to ever permanently quit smoking? (Circle one)

   A B C D E

ABOUT SMOKING

This affective measure assesses participants' intention to refrain from smoking cigarettes. This measure is appropriate for adolescents and preadolescents.

PURPOSE

Information about participants' intention to refrain from smoking may be useful for the following reasons:

- Administration of this measure at the beginning of the program may provide needs assessment information. For example, results of this measure may indicate that participants have little intention to refrain from smoking, thus emphasizing a need for instruction in that area.
- When this measure is administered prior to and following a program, the results will demonstrate the program's effects on participants' intention to refrain from smoking.

PROCEDURES

This instrument can be administered both at the beginning and at the end of the program. However, handbook users should be alert to concerns regarding the potential reactivity of affective measures. A measure is considered *reactive* if the experience of completing the measure prior to the program causes participants to react differently to the program. Handbook users should, therefore, carefully review each affective measure that they wish to use to determine its potential for making participants unduly sensitive to aspects of the program. If a measure is determined to be reactive, then program personnel should *not* administer that measure to *all* participants as a pretest and posttest. Instead, the measure could be administered to half of the program participants prior to program participation to determine participants' pre-program status. The measure could then be administered to the other half of the participants after program participation to assess participants' post-program status.

SCORING AND ANALYSIS

Point values are assigned to responses as follows:

Definitely Yes	=	1
Probably Yes	=	2
Maybe	=	3
Probably No	=	4
Definitely No	=	5

This inventory can be scored by adding the point values of the responses from all participants and dividing this total by the number of responses. Blank items should not be counted in the number of responses. The maximum attainable score of 5 points indicates a strong intention to refrain from smoking. A minimum score of 1 suggests little intention to refrain from smoking.

ABOUT SMOKING

The questions below are about whether you will smoke cigarettes in the future. Put a check to show your answer for each question.

	Definitely Yes	Probably Yes	Maybe	Probably No	Definitely No
1. Will you smoke any cigarettes during the next *month*?	( )	( )	( )	( )	( )
2. Will you smoke any cigarettes during the next *year*?	( )	( )	( )	( )	( )
3. When you are an adult, will you be a smoker?	( )	( )	( )	( )	( )

IDEAS ABOUT DECISIONS

This affective measure assesses participants' belief in the value of careful decision making. This measure is appropriate for adolescents and preadolescents.

PURPOSE

Information about decision making may be useful for the following reasons:

- Administration of this measure at the beginning of the program may provide needs assessment information. For example, results of this measure may indicate a need for strengthening participants' appreciation for careful decision making in dealing with smoking related situations in their lives.
- When this measure is administered prior to and following a program, it is possible to evaluate changes in participants' beliefs regarding careful decision making.

PROCEDURES

This instrument can be administered both at the beginning and at the end of the program. However, handbook users should be alert to concerns regarding the potential reactivity of affective measures. A measure is considered *reactive* if the experience of completing the measure prior to the program causes participants to react differently to the program. Handbook users should, therefore, carefully review each affective measure that they wish to use to determine its potential for making participants unduly sensitive to aspects of the program. If a measure is determined to be reactive, then program personnel should *not* administer that measure to *all* participants as a pretest and posttest. Instead, the measure could be administered to half of the program participants prior to program participation to determine participants' pre-program status. The measure could then be administered to the other half of the participants after program participation to assess participants' post-program status.

SCORING AND ANALYSIS

Point values are assigned to responses as follows:

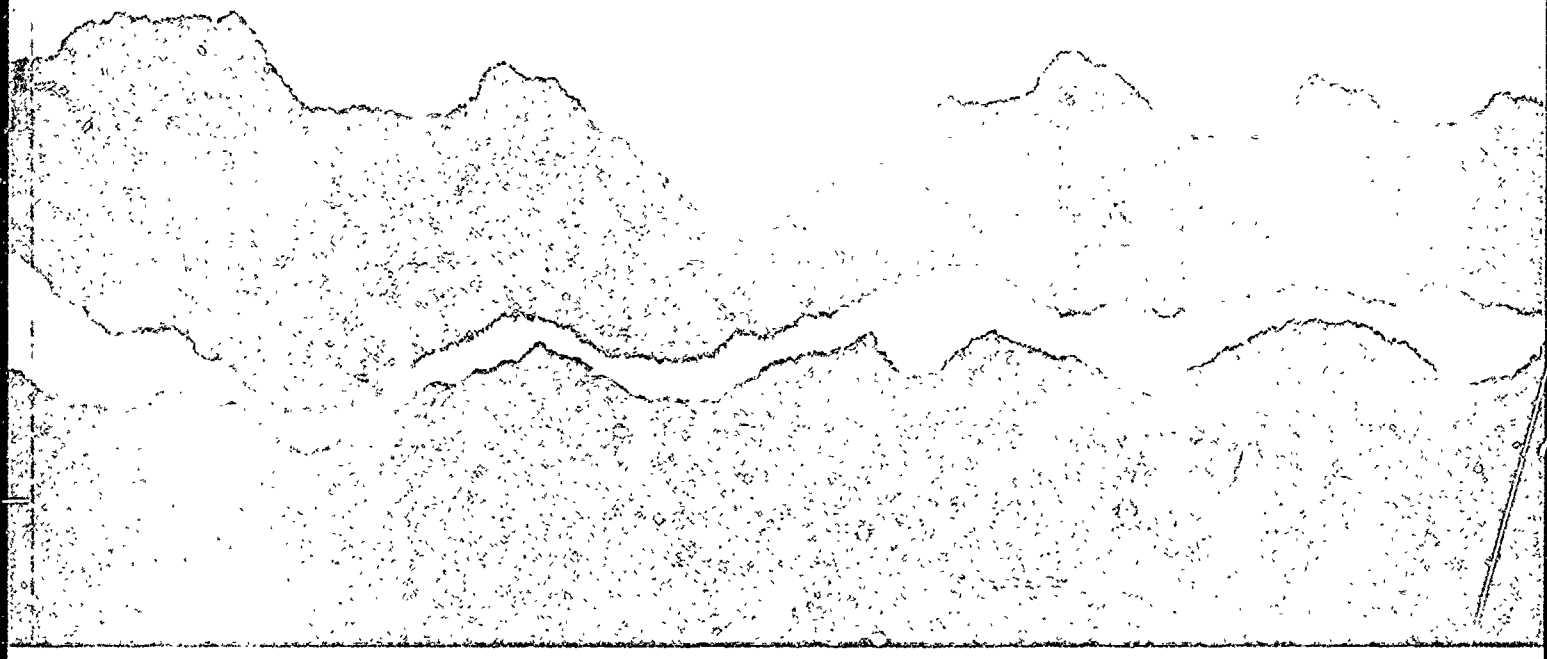
Item No.	Strongly Agree	Agree	Not Sure	Disagree	Strongly Disagree
1	5	4	3	2	1
2	1	2	3	4	5
3	5	4	3	2	1
4	1	2	3	4	5
5	1	2	3	4	5
6	1	2	3	4	5
7	1	2	3	4	5
8	5	4	3	2	1
9	5	4	3	2	1
10	5	4	3	2	1

This inventory can be scored by adding the point values of the responses from all participants and dividing this total by the number of responses. Blank items should *not* be counted in the number of responses. The maximum attainable score of 5 points indicates a strong belief in the utility of making decisions carefully.

IDEAS ABOUT DECISIONS

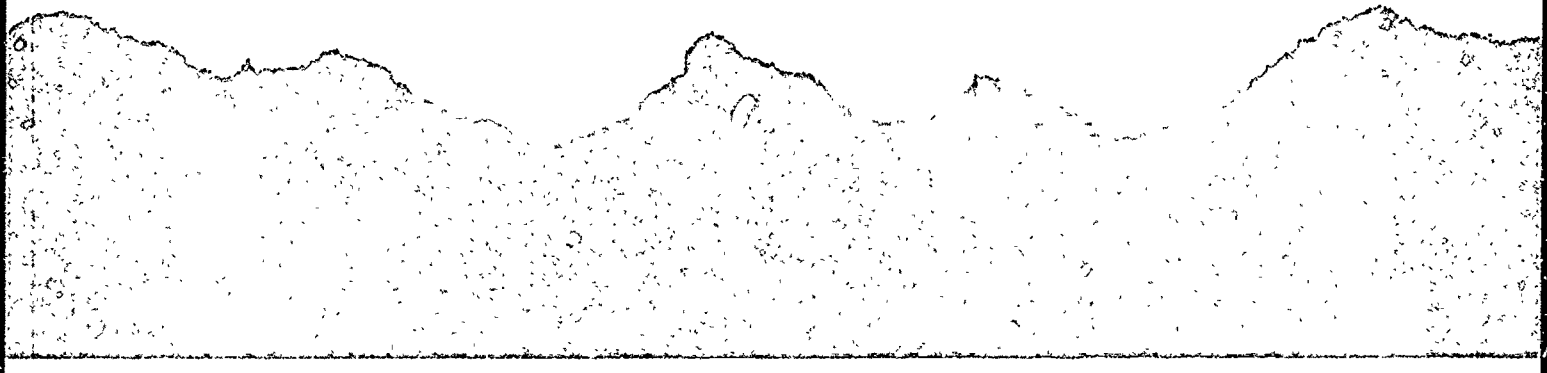
The sentences below are about making decisions. For each sentence, place a check to show how much you agree or disagree with the sentence.

	Strongly Agree	Agree	Not Sure	Disagree	Strongly Disagree
1. It is worth the time it takes to make decisions carefully.	( )	( )	( )	( )	( )
2. People should go with their first ideas when making decisions.	( )	( )	( )	( )	( )
3. People are happier with their decisions when they take the time to make them carefully.	( )	( )	( )	( )	( )
4. Spending a lot of time to make careful decisions is too difficult.	( )	( )	( )	( )	( )
5. Making careful decisions takes too much time.	( )	( )	( )	( )	( )
6. When making decisions, people should do what they feel, not what they think.	( )	( )	( )	( )	( )
7. People make equally good decisions no matter how they arrive at them.	( )	( )	( )	( )	( )
8. People who make quick decisions are usually disappointed with them later.	( )	( )	( )	( )	( )
9. People should take time to make decisions carefully.	( )	( )	( )	( )	( )
10. It is easy to make decisions carefully.	( )	( )	( )	( )	( )



CHAPTER FOUR

Locally Conducted Psychometric Studies



Locally Conducted Psychometric Studies

As described in Chapter One, the first step in using the newly developed handbook measures to examine program effectiveness is to select those measures that match program goals. However, evaluators cannot assume that a measure that appears to assess a desired program outcome will produce valid data about that outcome. When evaluators use a measure, they first want to determine the technical quality of that measure to ensure that any conclusions drawn about a program's effects are warranted. The purpose of this chapter is to assist evaluators in conducting validation studies for those handbook measures chosen for use in program evaluation.

Determining the Technical Quality of Measuring Devices

The degree to which a measuring instrument yields scores from which one can make legitimate inferences is referred to as validity. Tests are not valid or invalid. Rather, it is the inferences made, based on test results, that are valid or invalid. It is, therefore, technically accurate to focus on the *validity of score-based inferences* rather than the validity of a particular measuring device.

The concept of validity is highly dependent on the particular way in which a measuring instrument will be used. For example, a measure of the use of coping techniques to avoid smoking may permit a valid inference regarding the *number* of different techniques that program participants use, but may yield invalid inferences regarding the *frequency* with which participants use each technique. Furthermore, a test may yield valid inferences for a particular purpose with one population but invalid inferences for the same purpose with a different population. Thus, because validity varies on the basis of purpose and population, it is most appropriate to examine validity in the setting in which a measure will be used.

A second factor in determining the technical quality of a measurement instrument deals with the extent to which the instrument produces reliable, that is, consistent, results. Because the newly developed handbook measures have been subjected only to small-scale field tests, no reliability data are currently available. It is hoped that handbook users will conduct their own reliability studies and share those results with the Centers for Disease Control. In this way, results can be compiled over time and, subsequently, provided to handbook users. Procedures for evaluating the reliability of the handbook measures will be presented following a discussion of local validation approaches.

Categories of Validity Evidence

There are three major types of evidence regarding validity. These include content-related evidence of validity, criterion-related evidence of validity, and construct-related evidence of validity. The procedures for securing each type of validity evidence will be described below.

Content-related evidence of validity. Content-related evidence of validity involves the careful review of a measure's content by individuals identified as experts in the content area being assessed. This type of validity evidence is particularly important for measures designed to assess examinees' knowledge. To secure positive content-related validity, the

measure must include only those items that correspond to the content area being assessed and its items must address all important facets of that content area. The systematic, expertise-rooted procedures used to develop the handbook's instruments helped to ensure that appropriate content was built into the measures. Subsequent reviews by external experts confirmed that the measures are, indeed, focused on suitable content. These development procedures and the role of expert advisors in the project are described in the handbook's preface.

If there are questions regarding the suitability of the content in any of the handbook's measures, content-related validity can be examined by assembling a panel of experts who can judge the suitability of a measure's content for the specific program-evaluation purpose for which the measure is to be used. A panel of approximately 10 knowledgeable individuals can be asked to review the measuring instrument's items, one by one, and render independent yes/no judgments regarding the appropriateness of each item's content (in relationship to the inference that the program evaluators wish to make on the basis of the measure). In addition, panelists can be asked to determine whether any important content has been omitted from the measure. For example, if a knowledge measure such as **Smoking and Society** is being reviewed, panelists might be asked to first think of all the important facts about smoking's societal effects that program participants must know and then to indicate the percentage of those facts that are present in the measure being reviewed. This straightforward indication of a measure's content representativeness, when coupled with judgments regarding the content appropriateness of a measure's items, can yield important content-related evidence of validity for a measure.*

Criterion-related evidence of validity. Criterion-related evidence of validity requires that a measure be checked against an independent criterion. The independent criterion or standard should be one that the measure would be expected to predict. Criterion-related validity is most important for the handbook measures in the areas of behavior and intention. In the area of behavioral self-reports, for example, criterion-related validity would focus on the degree to which the self-reports reflect actual behavior. So, for example, criterion-related validity for a self-report instrument designed to measure the use of coping techniques would be secured by correlating responses on this instrument with observations (by others) of the extent to which the techniques were *actually* being used.

External criterion measures, such as observations, while often more accurate measures of behavior than self-reports, are extremely costly and time consuming to use. Thus, although it may be possible to use such criterion measures in a one-time validity study, they typically will not eliminate the need for self-report instruments in routine program evaluations. The general procedure for conducting a criterion-related validity study is shown in Figure 4.1.

A correlation of approximately .50 or higher between the measure and criterion would indicate that the new measure is predictive of the external criterion measure and, therefore,

---

* For additional information about how to conduct content related validation studies, see Annotated Bibliography Nos. 18, 23, 27, and 34.

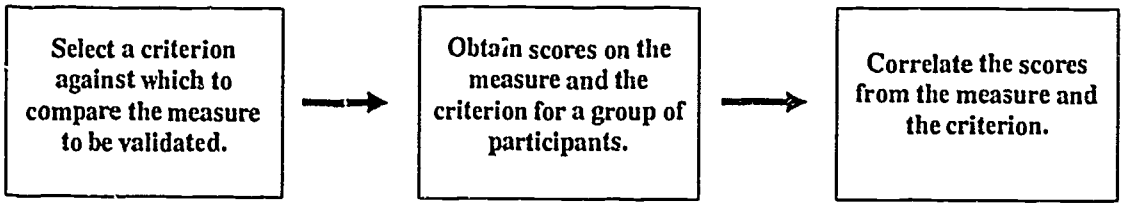


Figure 4.1: Procedure for conducting criterion-related validity studies

is measuring what it is intended to measure. A low correlation would call into question the self-report instrument as a measure of the behavior of interest.

Each criterion-related validity study must be specifically designed for the particular measure being examined and the purpose for which it will be used. For example, imagine that an evaluator wanted to examine the criterion-related evidence of validity for the handbook's measure entitled **Smoking Survey**. The evaluator must first identify an appropriate criterion measure. How is a program evaluator likely to use an intention measure? The most likely use would be to employ it as a proxy measure foreshadowing a program's effect on the future behavior of participants. That is, will program participants continue to refrain from smoking in the future? Thus, an appropriate criterion measure might be the reported smoking levels several months following the program.

To assemble criterion-related evidence of validity for the intention measure, a program evaluator could administer the intention measure at the end of the program to a group of at least 30 participants (or repeat this process each session until responses from at least 30 participants are obtained) and obtain completed self-report surveys several months later regarding participants' smoking levels. Once both measures are collected for every individual, a correlation could be computed between the strength of intention not to smoke and whether the participants smoked following the program. Thus, the criterion-related validity study would examine whether the intention measure was, in fact, predictive of later behavior. A measure that can serve as a meaningful proxy for participants' future behavior can prove highly useful in the evaluation of a program's impact on participants.*

Construct-related evidence of validity. The final type of validity evidence to be reviewed, construct-related evidence of validity, is particularly important for those handbook measures that do not have a clear criterion against which they can be evaluated. Such measures include the attitudinal and affective measures such as **Refraining from Smoking**, a measure that examines an individual's perceived ability to refrain from smoking in certain situations where people might want to smoke. Construct-related validity involves the gradual accumulation of data regarding what a test measures. Three strategies are customarily used to secure construct-related evidence of validity for a measure. First, in the *related-measures strategy,* predictions can be tested about the extent to which the measure of

---

* For additional information about the design and analysis of criterion-related validity studies, see Annotated Bibliography Nos. 18, 23, 27, and 34.

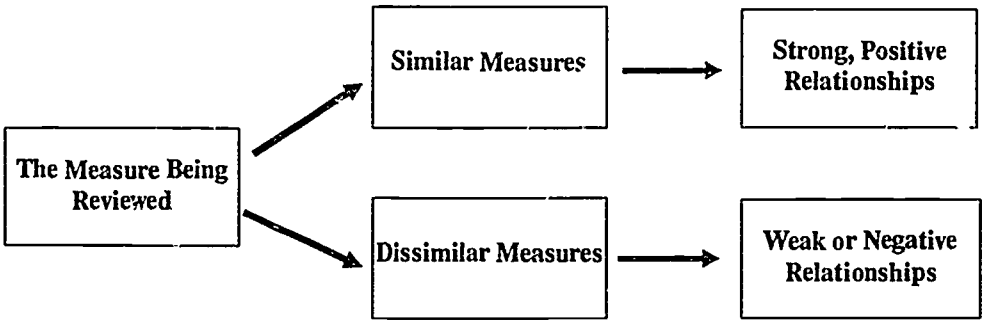


Figure 4.2: Correlations between measures assessing similar/dissimilar attitudinal dimensions

interest is correlated with other measures. For example, perceived ability to not smoke should be positively related to other measures aimed at assessing a similar attribute but should show reduced correlations with measures tapping different attitudinal dimensions. Thus, other existing measures can be correlated with the measure of interest to help clarify what is being measured.

If the correlations are consistent with the prior predictions, then construct-related evidence of validity has been obtained to support the defensibility of inferences based on the measure's use. Figure 4.2 illustrates the anticipated correlations between the measure of interest and other similar and dissimilar measures.

A second approach to examining construct-related validity involves predictions about group differences and is referred to as a *differential-populations strategy.* For this procedure, two or more groups are identified which are expected, based on other characteristics, to perform differently on the measure of interest. For example, the two groups might be individuals who have spouses or other family members who smoke versus those who do not. If the anticipated performance difference between the two groups is not obtained, it would raise the question as to whether the test was measuring what it was thought to measure.

A third strategy for securing construct-related evidence of validity is referred to as an *intervention strategy* because it involves the use of interventions such as training programs. For instance, a measure examined via this strategy could be administered to a group of participants before and after a "proven" smoking cessation program. If a difference in participants' scores on the measure is not observed, then the construct-related evidence of validity regarding the measure being reviewed is not supportive of the measure's use.

Construct-related evidence of validity is never based on a single study. Instead, consideration of a variety of studies, employing multiple validation strategies such as those described here, will help provide greater clarification regarding the appropriateness of using a given measuring instrument.*

Types of Reliability

A second characteristic of a defensible measurement instrument is the reliability or consistency with which it measures. The reliability of a test can be examined in three distinct ways. These include test-retest reliability, alternate-forms reliability, and internal consistency. Each of these approaches will be described below.

Test-retest reliability. Test-retest reliability (also referred to as *stability* reliability) examines the extent to which a measurement instrument is consistent over testing occasions. That is, will an individual who received a particular score on one testing occasion receive a similar score on a different testing occasion? Typically, to secure test-retest reliability information, an instrument is administered once to a group of individuals (30 or more). The same instrument is then administered again under similar conditions to the same group of individuals approximately two to four weeks later. Individuals' scores from the two administrations are then correlated. The higher the correlation, the greater the stability of measurement over time. Short tests, or other tests that are likely to be easily remembered, may result in an overestimate of reliability if participants recall their answers and, hence, respond similarly on the second testing occasion.

Alternate-forms reliability. The knowledge measures in this handbook have two forms that may be used for a pretest to posttest comparison. The administration of one form for the pretest and the other form for the posttest is desirable because the pretest may sensitize participants to pay more attention to those issues included on the pretest than to other equally important issues. However, to draw defensible conclusions based on the use of two different forms at pretest and posttest, the forms must be equivalent.

To examine alternate-forms reliability, it is necessary to administer both forms to the same group of individuals. The scores from the two forms can then be correlated. High correlations indicate that the same conclusions would be drawn about an individual or group of participants regardless of which of the two forms had been used. Thus, there would be reliable or consistent measurement across alternate forms. A high alternate-forms reliability coefficient does not guarantee that the forms are perfectly equidifficult. If the two forms are not of equal difficulty, that is, participants perform consistently better on one form than the other, it would still be possible to obtain high between-forms correlations. Thus, it is important to be attentive to mean scores on the two test forms. It is also permissible to use p-values (the percentage of examinees getting each item correct) to reassign items to forms so that they are more equidifficult. After the redistribution of items, a second alternate-forms reliability study should be conducted.

---

* For additional information about how to conduct construct-related validity studies, see Annotated Bibliography Nos. 18, 23, 27, and 34.

Handbook users should not assume equivalence or equidifficulty for the multiple forms provided in this handbook. Until alternate-forms reliability and test difficulty are examined, the measures should be used in a design such that half of the participants take Form A as a pretest and Form B as a posttest while the other half take Form B as a pretest and Form A as a posttest. This counterbalancing technique eliminates the possible influence of one form being more difficult than the other.

Internal consistency. Internal consistency examines the extent to which the instrument measures a single or related set of constructs. The higher the internal consistency, the greater the homogeneity of items on the test. A test thought to measure a single attitudinal dimension should have relatively high internal consistency reliability. Procedures for calculating internal consistency measures include split-half reliability, Kuder-Richardson formulas, and Cronbach's Alpha. The split-half reliability coefficient is calculated by administering the test to a group of at least 30 participants and then correlating scores from the odd versus the even items. A correction for test length must then be made using the Spearman-Brown formula. The split-half procedure is very similar to alternate-forms reliability in that two "forms" are correlated by separating the odd and even items. Kuder-Richardson formulas for internal consistency provide an estimate of the average of all possible split-halves. These formulas, like Spearman-Brown, require that test items be binary-scored, that is, able to be scored as right or wrong. Cronbach's Alpha is identical to Kuder-Richardson for binary scored items but can also be used for items that yield responses to which several points can be assigned, such as items on **Beliefs About Smoking.**

Not all forms of reliability need to be computed for every test. For example, alternate-forms reliability would be computed only for those measures that have two forms. Internal consistency estimates are less appropriate for multidimensional measures. Test-retest reliability is appropriate for most measures, but often presents pragmatic problems due to the need to retest the same individuals.*

Groups and Individuals

The validity and reliability procedures reviewed here were originally developed to examine the quality of tests used for *individual* assessment purposes. In contrast, the recommended use of the handbook measures is to perform *group* analyses for program evaluation. Thus, the appropriate reliability issue is whether scores for a group of individuals are relatively consistent. Similarly, the validity issue is whether changes in scores for a group of individuals are reflective of changes in the group's knowledge, affect, or behavior. Because group scores are more stable than individual scores, the procedures outlined above are likely to underestimate the reliability and validity of the measures when used for program evaluation. Practically speaking, a measurement instrument with a lower reliability or validity coefficient would be acceptable when used for group rather than individual

---

* For additional information about how to examine the reliability of measurement instruments, see Annotated Bibliography Nos. 3, 18, 19, 23, 27, and 34.

diagnosis. For example, Salvia and Ysseldyke (1981, p 98) have recommended the following minimum standards for alternate-forms reliability:

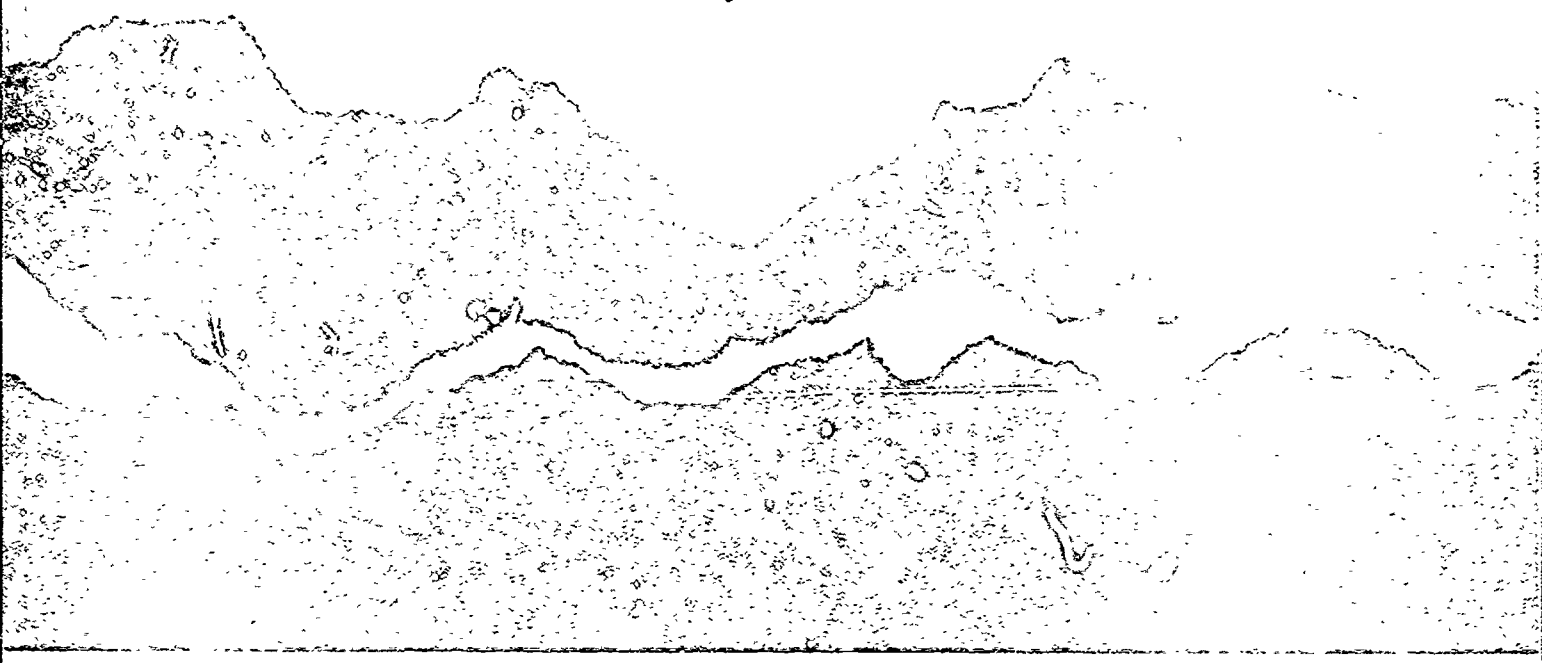
.60 - when scores are reported for groups

.80 - when scores are used for individual screening

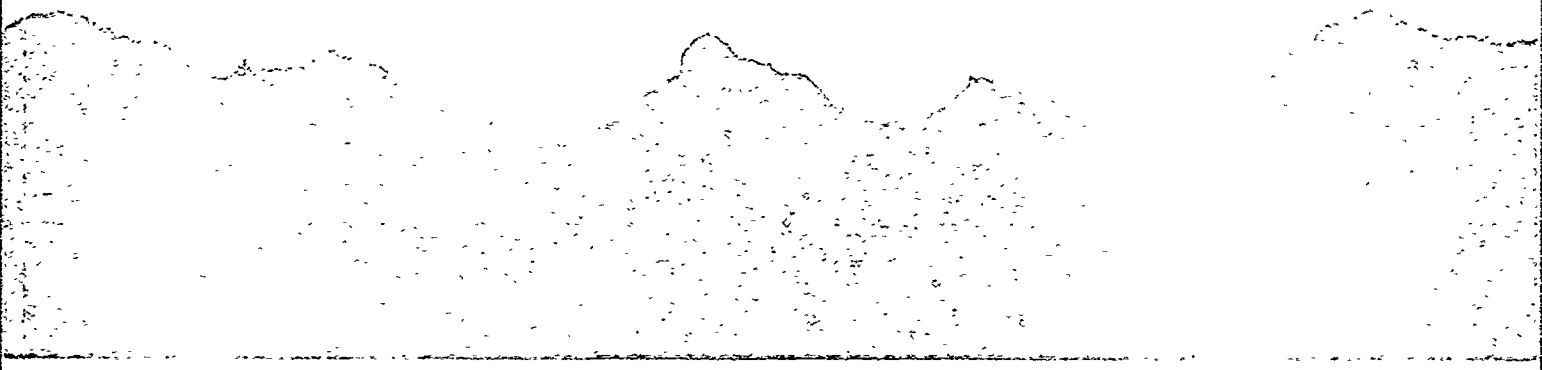
.90 - when scores are used for important educational decisions for individuals

Thus, standards for acceptable reliability and validity vary depending on the purpose for using a particular measure. However, minimal levels for each are critical for making sound decisions about a program. With a little creativity and effort, studies of reliability and validity can often be integrated into the ongoing operation of a program.

In addition to providing a brief overview, the major purpose of this chapter was to encourage handbook users to conduct local reliability and validity studies and to consider the involvement of a measurement specialist or the use of appropriate references in designing such studies. As suggested at the outset of the chapter, if such local studies are carried out, results should be forwarded to the Centers for Disease Control (Attention: Dr. Diane Orenstein, Project Officer, Center for Health Promotion and Education, Centers for Disease Control, 1600 Clifton Road N.E., Atlanta, GA 30333). This information will be shared with future handbook users.



Appendices



Appendix A

AMPLIFIED CONTENT DESCRIPTORS*

THE PHYSICAL EFFECTS OF SMOKING

(Adult/Adolescent Measure)

FACTS ABOUT SMOKING

(Adolescent/Preadolescent Measure)

General Biomedical Consequences of Smoking

1. Tobacco smoke consists of dangerous particles and gases.
2. Tar, nicotine, hydrogen cyanide, and carbon monoxide are inhaled when you smoke.
3. Tar consists of numerous chemicals, some of which are believed to cause cancer.
4. Some scientists believe that nicotine is addictive.
5. Nicotine causes blood vessels to decrease in size, which reduces the amount of blood that can be transported.
6. Nicotine causes the heart to beat more rapidly.
7. Nicotine produces drug-like dependence in smokers.
8. Hydrogen cyanide damages the respiratory system.
9. Carbon monoxide decreases the amount of oxygen in the blood.

Smoking and Disease

10. The risk of developing coronary heart disease is twice as great for cigarette smokers as for nonsmokers.
11. The risk of developing coronary heart disease increases with the number of cigarettes smoked.
12. Cigarette smoking is directly related to one in every three deaths from cancer.

---

* The amplified content descriptors are not exhaustive accounts of smoking cessation content. At the time this document was prepared, the most current statistical information available had been gathered in 1985-86 and was published between 1986 and 1988. You may be able to update these descriptors by referring to more recent editions of the documents cited in the bibliography.

13. People who smoke frequently have trouble breathing and usually cough a lot.
14. Cigarette smoking increases the risk of developing cancer of the lung, larynx, pharynx, mouth, esophagus, kidney, pancreas, and bladder.
15. People who smoke have more gum and mouth problems than people who do not smoke.
16. The risk of developing lung cancer is ten times greater for cigarette smokers than for nonsmokers.
17. The risk of developing lung cancer increases proportionately with the number of cigarettes smoked each day, the number of years of smoking, and the depth to which the cigarette smoke is inhaled.
18. Besides having illnesses such as cancer and heart problems, people who smoke are usually sick more often than people who do not smoke.
19. Cigarette smoking increases the risk of developing lung cancer for both men and women.
20. Cigarette smoking increases the risk of developing chronic bronchitis and emphysema.
21. People who smoke are likely to die at a younger age than people who do not smoke.
22. The number of people who annually die prematurely from smoking-related diseases is estimated to be over 300,000.
23. Even small levels of smoking can be bad for a person's health.
24. Smoking can stain a person's teeth.
25. Smoking can leave a bad smell on a person's breath and clothing.

Interactive Effects of Smoking

26. Cigarette smoking increases the risk of being harmed by exposure to other dangerous materials such as asbestos or coal dust.
27. Cigarette smoking increases the dangers associated with taking birth control pills.
28. Smokers have an increased risk of developing a respiratory infection after an operation.
29. Cigarette smoking during the later months of pregnancy increases the risk of having a stillborn baby, a baby that dies shortly after birth, or a baby of lower than average birthweight.
30. If a woman smokes during pregnancy, the nicotine and carbon monoxide she inhales enter the blood of the fetus.
31. Cigarette smoking during pregnancy increases the risk of having a baby who will get "sudden infant death syndrome."

Effects of Smoking Filter Cigarettes

32. Smokers of filter cigarettes are four times more likely than nonsmokers to develop lung cancer.
33. Smokers of filter cigarettes are at less risk of developing lung cancer than smokers of nonfilter cigarettes.
34. Smokers of filter cigarettes are at less risk of developing respiratory diseases than smokers of nonfilter cigarettes.
35. Smokers of most filter cigarettes inhale more carbon monoxide than smokers of nonfilter cigarettes.
36. Smokers of most filter cigarettes are probably at greater risk of developing coronary heart disease than smokers of nonfilter cigarettes.

Effects of Smoking Low-Tar and Low-Nicotine Cigarettes

37. Death rates are lower for smokers of low-tar and low-nicotine cigarettes than for smokers of high-tar and high-nicotine cigarettes.
38. Death rates are higher for smokers of low-tar and low-nicotine cigarettes than for nonsmokers.
39. Many smokers inhale more deeply when they smoke low-tar and low-nicotine cigarettes, offsetting the reduced health risks.

Effects of Smoking Pipes and Cigars

40. Pipe or cigar smokers are less likely than cigarette smokers to develop lung cancer.
41. Pipe or cigar smokers are more likely than nonsmokers to develop lung cancer.
42. Pipe or cigar smokers who inhale while they are smoking are at greater risk of developing lung cancer than pipe or cigar smokers who do not inhale.
43. Pipe or cigar smokers are at the same risk as cigarette smokers of developing cancer of the esophagus, pharynx, larynx, and mouth.
44. Pipe smoking increases the risk of developing lip cancer.

Effects of Quitting Smoking

45. Most regular smokers may feel nervous and shaky when they first stop smoking.
46. The health risks associated with smoking decrease when a person stops smoking.
47. If a person quits smoking for 10-15 years, that person's chances of developing lung cancer are the same as a nonsmoker's chances.
48. The same number of people lose weight as gain weight after giving up smoking.

49. If a woman stops smoking by the fourth month of her pregnancy, the risks of health problems or death to her infant are probably reduced to the levels of those for nonsmoking women.

50. Most smokers say they want to quit and that they have made at least one serious attempt to do so.

Effects of Involuntary Smoking

51. Smokers and nonsmokers can suffer eye irritation, headaches, and nose and throat discomfort from cigarette smoke

52. Cigarette smoke may fill an enclosed area with higher levels of carbon monoxide and other pollutants than are usually present during an air pollution emergency.

53. Infants whose parents smoke have a greater chance of developing respiratory infections than do infants whose parents do not smoke.

54. Parents who smoke are more likely to have children who smoke than are parents who do not smoke.

SMOKING AND SOCIETY

(Adult/Adolescent Measure)

PROBLEMS WITH SMOKING

(Adolescent/Preadolescent Measure)

Economic Costs in the United States

1. The health consequences of smoking are estimated to cost over $30 billion annually.
2. The effects of smoking cost society more than either the cost of drug abuse or alcohol abuse.
3. For every dollar spent purchasing cigarettes, smokers spend at least another dollar directly on health care costs for smoking-related diseases.
4. Americans annually spend over 30 billion dollars purchasing cigarettes.
5. The estimated cost of lost earnings due to sickness and death because of cigarette smoking is $50 billion.
6. Over 10% of all United States direct health care costs are attributable to cigarette smoking.

Effects of Involuntary Smoke

7. Children of smokers have greater risk of hospitalization for bronchitis and pneumonia than do children of nonsmokers.
8. Involuntary or environmental tobacco smoke can cause lung cancer in healthy nonsmokers.
9. Nonsmokers can suffer eye irritation, headaches, and nose and throat discomfort from cigarette smoke.
10. Simple separation of smokers and nonsmokers within the same air space does not eliminate exposure of nonsmokers to environmental tobacco smoke.
11. Effective removal of smoke particles from indoor air requires an increase in the exchange with outdoor air.
12. Greater amounts of carcinogens are found in sidestream smoke than in mainstream smoke.

Smoking and Premature Death in the American Population

13. Approximately 300,000 people annually are estimated to die prematurely from the effects of smoking.
14. Of the 60,000 people who die each year from chronic obstructive lung disease, about 85% of the deaths can be directly attributed to smoking.

15. Smoking-related cardiovascular disease accounts for the most deaths of any preventable disease.
16. Of the 135,000 lung cancer deaths per year, 85% are directly attributable to cigarette smoking.
17. The coronary heart disease death rate is 70% higher for smokers than it is for nonsmokers.
18. About 15% of all deaths are related to smoking.
19. Smokers' risk of death from lung cancer is ten times greater than that of nonsmokers.
20. Cigarettes start fires that account for 30% of all fire-related deaths.

Societal Issues

21. Women find it more difficult to quit smoking than men do.
22. Maternal smoking contributes to prenatal mortality and low birth weight.
23. The damaging effects on fetal development are reduced if the mother stops smoking during her pregnancy.
24. Approximately 25% of the adult population currently smoke.
25. In families where both parents smoke, about 25% of the adolescent children also smoke.
26. The current trend in teenage smoking is for the number of boys who smoke to stay the same while the number of girls who smoke increases.
27. The primary reasons that teenagers begin smoking appear to be peer group pressure and other social influences.
28. Even though the gap between the number of men who smoke and women who smoke is narrowing, currently more men smoke than women.
29. The percentage of smokers who are quitting has gradually increased since the Surgeon General reports on the health risks of smoking were first published in the 1960s.
30. The number of children who smoke is two times higher for children of two smoking parents than it is for children of two nonsmoking parents.
31. The general trend for children who smoke is that children who smoke are becomingregular smokers at a younger age.
32. Only about 5% of adolescents age 12-14 smoke once a week or more.
33. Children who experiment with smoking grow up to be regular smokers more often than children who do not experiment.

Smoking Cessation Content Bibliography

American Cancer Society. (1972). *The dangers of smoking; the benefits of quitting.* New York: Author.

American Lung Association. (1980). *A lifetime of freedom from smoking.* New York: Author.

Bell, C.S., & Grabowski, J. (Eds.). (1983). *Measurement in the analysis and treatment of smoking behavior* (Research Monograph No. 48). Rockville, MD: National Institute on Drug Abuse.

Coelho, R.J. (1985). *Quitting smoking: A psychological experiment using community research.* New York: Peter Lang.

Danaher, B.G., & Lichtenstein, E. (1978). *Become an ex-smoker.* Englewood Cliffs, NJ: Prentice-Hall.

Holslcaw, D.S., & Topham, A.L. (1978). The effects of smoking on fetal, neonatal, and childhood development. *Pediatric Annals, 7*(3), 105-136.

Jarvik, M.E., Cullen, J.W., Gritz, E.R., Vogt, T.M., & West, L.J. (Eds.). (1977, December). *Research on smoking behavior* (Research Monograph No. 17). Rockville, MD: National Institute on Drug Abuse.

Krasnegor, N.A. (Ed.). (1979, January). *Cigarette smoking as a dependence process* (Research Monograph No. 23). Rockville, MD: National Institute on Drug Abuse.

Krasnegor, N.A. (Ed.). (1979, August). *The behavioral aspects of smoking* (Research Monograph No. 26). Rockville, MD: National Institute on Drug Abuse.

Lachter, S.F., & Weisman, R. (1980). *Let's talk about drug abuse.* Rockville, MD: U.S. Department of Health and Human Services.

National Cancer Institute. (1977). *The smoking digest: Progress report on a nation kicking the habit.* Bethesda, MD: Author.

Smith, S.C., Levit, A.B., & Pinkston, K.B. (1981). *Catching on.* Rockville, MD: U.S. Department of Health and Human Services.

Sonnett, S. (1977). *Smoking.* New York: Franklin Watts.

U.S. Department of Health, Education, and Welfare. (1979). *Smoking and health: A report of the surgeon general.* Washington, DC: Author.

U.S. Department of Health, Education, and Welfare Public Health Service, Centers for Disease Control. (1975). *The health consequences of smoking.* Washington, DC: Author.

U.S. Department of Health and Human Services. (1980). *The health consequences of smoking for women: A report of the surgeon general.* Rockville, MD: Author.

U.S. Department of Health and Human Services. (1982). *The health consequences of smoking: Cancer. A report of the surgeon general.* Rockville, MD: Author.

U.S. Department of Health and Human Services. (1983). *The health consequences of smoking: Cardiovascular disease. A report of the surgeon general.* Rockville, MD: Author.

U.S. Department of Health and Human Services. (1984). *The health consequences of smoking: Chronic obstructive lung disease. A report of the surgeon general.* Rockville, MD: Author.

U.S. Department of Health and Human Services. (1986). *The health consequences of involuntary smoking: A report of the surgeon general.* Rockville, MD: Author.

U.S. Department of Health and Human Services. (1988). *The health consequences of smoking: Nicotine addiction. A report of the surgeon general.* Rockville, MD: Author.

Appendix B

INFORMED CONSENT PROCEDURES

Prior to administering measures to participants, program personnel should inform participants about the content covered by the measures and the purpose of the program's evaluation study. Program personnel may also wish to provide the opportunity for participants to indicate whether or not they consent to participate in the study and complete the selected measures. Informed consent is obtained by presenting all information pertinent to the study and asking the participant to affix a signature indicating that the information has been read and that consent is given to participate.

If the decision is made to obtain informed consent, program personnel have the choice of employing a "passive" consent procedure or an "active" consent procedure. *Passive informed consent* consists of asking participants to sign and return a consent form only *if they do not wish to participate* in the study. Participants who do not return the consent form are considered eligible to participate in the study.

Active informed consent requires participants to sign and return the consent form *if they wish to participate*. Only those participants who return a signed form can be included in the study. Consequently, the participation rate resulting from an active consent procedure is generally lower than that obtained from a passive consent procedure.

To construct an informed consent form, program personnel should consider including the following items:

1. A general statement of the program goals and objectives.
2. A brief explanation of the study procedures and measures.
3. An indication that the participant is free to withdraw consent and to discontinue participation at any time.
4. An explanation of the procedures to be taken to ensure anonymity and confidentiality of responses.
5. An indication that participants are free not to answer specific items or questions.
6. A place for the participants to affix their signatures under a statement indicating that the participant agrees to participate (active consent) or does not agree to participate (passive consent) in the study. If appropriate, a date for the return of the consent form should be specified.

Appendix C

ANNOTATED EVALUATION BIBLIOGRAPHY

1. Alkin, M.C., & Solmon, L.C. (Eds.). (1983). *The costs of evaluation*. Beverly Hills, CA: Sage.

   In this collection of essays both theoretical and practical issues relevant to cost-focused program evaluations are presented.

2. American Psychological Association. (1973). *Ethical principles in the conduct of research with human participants*. Washington, DC: Author.

   This treatise focuses on the appropriateness of carrying out various types of research investigations with human subjects. Because the American Psychological Association has had a long-standing concern about ethical issues in the conduct of research investigations, this publication will be of interest to numerous evaluators of health education programs.

3. American Psychological Association, American Educational Research Association, National Council on Measurement in Education. (1985). *Standards for educational and psychological tests*. Washington, DC: Author.

   This volume presents the most widely used set of standards for psychological and educational tests. Frequently cited by users of educational tests, the standards have recently been employed in numerous judicial deliberations. Relatively brief, the standards should be consulted by health educators who employ assessment devices regularly.

4. Anderson, L.W. (1981). *Assessing affective characteristics in the schools*. Boston: Allyn and Bacon.

   Anderson provides an excellent set of practical suggestions for the creation of affective assessment instruments. He includes one of the most easily understood expositions of various scaling procedures including Likert, Thurstone, and Guttman scales.

5. Bausell, R.B. (Ed.). *Evaluation and the health professions*. Newbury Park, CA: Sage.

   This quarterly publication deals with a variety of evaluation-relevant issues of interest to health educators.

6. Berk, R.A. (Ed.). (1982). *Handbook of methods for detecting test bias*. Baltimore: The Johns Hopkins University Press.

   This collection of individual essays offers the reader a comprehensive depiction of methods currently available to detect the presence of bias in tests.

7. Berk, R.A. (Ed.). (1984). *A guide to criterion-referenced test construction*. Baltimore: The Johns Hopkins University Press.

This collection of essays consists of papers presented at the first Johns Hopkins University National Symposium on Educational Research. In addition, a number of more recently written chapters have been included in this revision of a 1980 text. The authors address many of the important problems, both conceptual and technical, facing developers and users of criterion-referenced measures.

8. Campbell, D.T., & Stanley, J.C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

This volume, originally a chapter in a larger volume, has had substantial impact on the fields of research and evaluation. Evaluators of health education programs will wish to consider this truly classic treatment of data-gathering designs suitable for experimental and quasi-experimental settings.

9. Churchill, G.A., Jr. (1979). *Marketing research: Methodclogical foundations* (2nd ed.). Hinsdale, IL: The Dryden Press.

Although written in the context of marketing research, this textbook covers several topics of vital importance in evaluation. Topics such as research design, data collection, sampling, and data analysis are covered in a readily understandable yet accurate way. An excellent resource.

10. Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.

Cohen offers a useful treatment of factors which should be considered when one draws samples for use in research or evaluation activities. Of special interest is the set of easy-to-use guidelines he offers for determining the estimated sample size necessary to detect differences between groups.

11. Cook, T.D., & Campbell, D.T. (1976). The design and conduct of quasi-experiments and true experiments in field settings. In M.D. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Chicago: Rand McNally.

This is an updated version of the famous exposition of quasi-experimental and experimental data-gathering designs by Donald T. Campbell and Julian C. Stanley (see Reference No. 8). An excellent discussion of four types of validity is featured in this essay.

12. Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

This widely cited volume provides a comprehensive treatment of quasi-experimental investigations in settings of substantial relevance to the concerns of health educators. There are excellent discussions of internal and external validity, including the various threats to both types of validity. A systematic consideration of the commonly used data-gathering designs is offered, including an extended appraisal of interrupted time-series designs.

13. Cordray, D.S., Bloom, H.S., & Light, R.J. (Eds.). (1987, Summer). *Evaluation practice in review* (New Directions for Program Evaluation, No. 34). San Francisco: Jossey-Bass.

This volume contains a set of thought-provoking chapters dealing with what has been learned about the practice of evaluation during the past decade. The chapters on evaluation politics by Eleanor Chelimsky and on naturalistic evaluation by Egon Guba would be of particular interest to evaluators of health education programs.

14. Cronbach, L.J. (1963). Course improvement through evaluation. *Teachers College Record, 64*, 672-683.

This article is an early piece, presenting the virtues of what would later be termed "formative" evaluation. It rings as true today as it did more than two decades ago, and it applies as much to evaluation in health education as it does to more traditional evaluation. Emphasizing the role of evaluation in gathering information that can improve programs, this article is well worth reading.

15. Cronbach, L.J. (1977). *Analysis of covariance in nonrandomized experiments: Parameters affecting bias.* Unpublished occasional paper, Stanford Evaluation Consortium, Stanford University.

A highly technical piece on the complications associated with using analysis of covariance, this article is recommended only for those prepared to handle a critical data-analysis problem in a sophisticated way.

16. Cronbach, L.J., Ambron, S.R., Dornbusch, S.M., Hess, R.D., Hornik, R.C., Phillips, D.C., Walker, D.F., & Weiner, S.S. (1980). *Toward reform of program evaluation.* San Francisco: Jossey-Bass.

This important book considers the function of evaluation in a pluralistic society and presents 95 theses on the role of evaluators and evaluations. In addition to providing a contemporary conception of evaluation, it provides a historical and multidisciplinary perspective of the field. This volume will be of considerable interest to those evaluating health education programs.

17. Cronbach, L.J., & Furby, L. (1970). How should we measure 'change' — or should we? *Psychological Bulletin, 74*, 68-80.

A technical treatise on the dangers associated with using gain scores. A very significant piece, but recommended only for those with some psychometric training.

18. Cunningham, G.K. (1986). *Educational and psychological measurement.* New York: Macmillan.

This is a standard introductory text focusing on the major topics associated with measurement as it applies to such tasks as program evaluation.

19. Ebel, R.L. (1979). *Essentials of educational measurement* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

This is a standard, easily read introductory text, covering important topics in the field of educational testing. Ebel, a prominent leader of traditional educational testing practices, provides a lucid treatment of a wide range of measurement topics.

20. Fetterman, D.M., & Pitman, M.A. (Eds.). (1986). *Educational evaluation: Ethnography in theory, practice, and politics.* Beverly Hills, CA: Sage.

This collection of essays touches on ethnographically oriented evaluation of educational programs. Health educators wishing to learn about this recently emphasized approach to educational evaluation will find this volume of interest.

21. Green, L.W. (1979). Research methods translatable to the practice setting: From rigor to reality and back. In S.J. Cohen (Ed.), *New directions in patient compliance* (pp.141-151). Lexington, MA: Lexington Books.

Green attends to a practical dilemma facing those who evaluate health education programs, namely, the necessity to make trade-offs between validity and feasibility in field settings. Six strategies for coping with evaluation under adverse circumstances are described.

22. Green, L.W., & Figa-Talamanca, I. (1974). Suggested designs for evaluation of patient education programs. *Health Education Monographs, 2* (1), 54-71.

In this essay Green and Figa-Talamanca suggest data-gathering designs for conducting evaluations of patient education programs. The authors also explore several issues related to evaluations of this variety.

23. Green, L.W., & Lewis, F.M. (1986). *Measurement and evaluation in health education and health promotion.* Palo Alto, CA: Mayfield.

This volume is an excellent resource for health educators concerned with the evaluation of their programs. Green and Lewis provide a series of useful explanations of topics in both measurement and health evaluation. Their expositions are peppered with practical examples drawn from health education and health promotion.

24. Hambleton, R.K., Swaminathan, H., Algina, J., & Coulson, D.B. (1978). Criterion-referenced testing and measurement: A review of technical issues and development. *Review of Educational Research, 48* (1), 1-48.

This is a comprehensive review of the field of criterion-referenced testing. Hambleton and his colleagues do a masterful job of isolating the key issues in criterion-referenced testing and describing results of research investigations bearing on those issues. Somewhat technical at times, this review is one of the more widely cited essays dealing with criterion-referenced testing.

25. Hays, W.L. (1973). *Statistics for the social sciences.* New York: Holt, Rinehart, and Winston.

This comprehensive text handles basic and advanced statistical considerations. Somewhat technical at points, Hays nonetheless provides an excellent set of step-by-step guidelines to statistical practice.

26. Joint Committee on Standards for Educational Evaluation. (1981). *Standards for evaluations of educational programs, projects, and materials*. New York: McGraw-Hill.

The development of these evaluation standards was spearheaded by a joint committee of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. Thirty standards are presented, addressing issues related to deciding whether to evaluate, defining the evaluation problem, designing the evaluation, budgeting for the evaluation, collecting and analyzing data, and reporting the evaluation. Intended for both consumers of evaluation and individuals conducting evaluations, this reference may be of most use to evaluators who are relatively new to the field.

27. Kubiszyn, T., & Borich, G. (1987). *Educational testing and measurement: Classroom application and practice* (2nd ed.). Glenview, IL: Scott-Foresman.

Another introductory text dealing with the nuts and bolts of measurement, this book will provide health educators with a good overview of educational measurement.

28. Levin, H.M. (1975). Cost-effectiveness analysis in evaluation research. In M. Guttentag & E.L. Struening (Eds.), *Handbook of evaluation research* (Vol. 2, pp. 89-122). Beverly Hills, CA: Sage.

This essay probes the important considerations involved in determining cost-effectiveness of programs in the context of educational evaluations. Theoretical as well as practical guidelines are provided.

29. Levin, H.M. (1983). *Cost-effectiveness: A primer* (New Perspectives in Evaluation, Vol. 4). Beverly Hills, CA: Sage.

This text is a splendid introduction to the fundamental concepts of cost analysis on program evaluation. Levin provides succinct descriptions along with advantages and disadvantages for cost-feasibility, cost-effectiveness, cost-benefit, and cost-utility analyses.

30. Linn, R.L., & Slinde, J.A. (1977). The determination of the significance of change between pre- and posttesting periods. *Review of Educational Research, 47*, 121-150.

This article reviews many of the major issues in the measurement of change from pretesting to posttesting periods and suggests possible alternatives. These authors share the general sentiment of many others in the field that "more is expected from gain scores than they can reasonably be expected to provide."

31. Lord, F.H. (1963). Elementary models for measuring change. In C.W. Harris (Ed.), *Problems in measuring change* (pp. 21-38). Madison: Wisconsin Press.

This is an early treatise on the problems associated with measuring change. Although this chapter rapidly becomes very technical, the early sections provide an intuitive explanation of the difficulties with using gain scores.

32. Mark, M.M., & Shotland, R.L. (Eds.). (1987, Fall). *Multiple methods in program evaluation* (New Directions for Program Evaluation, No. 35). San Francisco: Jossey-Bass.

Decrying the infrequency with which multiple methods are used in program evaluation, six chapters are offered in this volume, not only advocating multiple methods, but also describing how such program evaluations can be conducted.

33. Oakland, T. (Ed.). (1977). *Psychological and educational assessment of minority children.* New York: Brunner/Mazel.

This collection of essays provides a series of useful suggestions for those who are more sensitive to the possible bias present in educational tests.

34. Popham, W.J. (1981). *Modern educational measurement.* Englewood Cliffs, NJ: Prentice-Hall.

Varied topics in the field of educational measurement are introduced in this text. Norm-referenced measurement and criterion-referenced measurement are both considered, with the special applications of criterion-referenced assessment emphasized. Chapters on the relationship of testing to teaching and the measurement of affect will be of special interest to health educators.

35. Popham, W.J. (1988). *Educational evaluation.* Englewood Cliffs, NJ: Prentice-Hall.

This is an introductory text, written in fairly nontechnical language, about the field of educational evaluation. Evaluators of health education programs will find it simple to translate the book's contents to their own specialties.

36. Popham, W.J., & Sirotnik, K.A. (1973). *Educational statistics: Use and interpretation* (2nd ed.). New York: Harper and Row.

This easily read introductory text deals with the fundamental types of statistical considerations needed by program evaluators. It is intended for those who are not particularly comfortable with mathematical approaches to statistics.

37. Riecken, H.W., & Boruch, R.F. (1971). *Social experimentation: A method for planning and evaluating social intervention.* New York: Academic Press.

This is a significant contribution to our thinking about large-scale social interventions, their design and appraisal. It provides a useful analysis of the ways that the experimental method can be defensibly employed in connection with major social programs.

38. Rivlin, A.M., & Timpane, P.M. (Eds.). (1975). *Ethical and legal issues in social experimentation.* Washington, DC: Brookings Institution.

Rivlin and Timpane explore the sorts of legal and ethical issues to which evaluators of health education programs must attend.

39. *SPSS-X User's Guide* (3rd ed.). (1988). Chicago: SPSS Inc.

This is a widely used, well-organized set of "canned" computer analysis programs for use in the social sciences. Health educators who have occasion to use computer analyses will find the SPSS manual most helpful.

40. Salvia, J., & Ysseldyke, J.E. (1981). *Assessment in special and remedial education* (2nd ed.). Boston: Houghton Mifflin.

This text, intended for individuals who must apply assessment to special education and remedial education, provides measurement insights for health educators who deal with such populations of learners.

41. Scriven, M. (1967). The methodology of evaluation. In R.W. Tyler, R.M. Gagné, & M. Scriven (Eds.). *Perspectives of curriculum evaluation* (pp. 39-83). Chicago: Rand McNally.

This seminal article was the first essay in which Scriven distinguished between the now commonly accepted formative and summative roles of evaluators. Scriven addresses a wide variety of topics, emphasizing the importance of comparative appraisals of two or more programs' merits.

42. Scriven, M. (1972). Prose and cons about goal-free evaluation. *Evaluation Comment, 3*, 1-4.

In this essay Scriven offers goal-free evaluation as an antidote to excessive preoccupation with the program staff's expressed objectives. Scriven argues that evaluators should attend to the results produced by a program, not the rhetoric of its program goals.

43. Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill.

This is the classic treatment of nonparametric statistical techniques. Although a bit out of date these days, Siegel's text offers the most easily understood treatment of nonparametric statistical procedures. Because of the author's admitted zealousness in support of nonparametric techniques, those using Siegel's text should also consult a critique of it by Robert Savage, *Journal of American Statistical Association*, 1957, *52*, 331-344.

44. Suchman, E.A. (1967). *Evaluative research: Principles and practice in public service and social action programs.* New York: Russell Sage Foundation.

In this volume, Suchman provides extensive coverage of the application of the experimental research model in conducting evaluations. Although evaluation has come a long way since this book was written, the volume provides a clear description of the predominant conceptualization of evaluation in the past decade.

45. Tukey, J.W. (1977). *Exploratory data analyses.* Reading, MA: Addison-Wesley.

Creative approaches to displaying and understanding data are provided by Tukey in this excellent demystification of data analysis.

46. Walberg, H.J., Postlethwaite, T.N., Creemers, B.P.M., & de Court, E. (Eds.). (1987). Educational evaluation: The state of the field. *International Journal of Educational Research, 11* (1).

This special issue, as its title suggests, presents comprehensive review of field of program evaluation from authors based in the U.S. and abroad.

47. Webb, E.J., Campbell, D.T., Schwartz, R.D., Sechrest, L., & Grove, J.B. (1981). *Nonreactive measures in the social sciences* (2nd ed.). Dallas: Houghton Mifflin.

This charming volume provides readers with a series of powerful and clever tactics to secure data, particularly of an affective nature, without sensitizing respondents to the evaluator's purposes.

48. Weiss, C.H. (1972). *Evaluation research: Methods of assessing program effectiveness.* Englewood Cliffs, NJ: Prentice-Hall.

Weiss offers a pithy overview of prominent program evaluation considerations including the formulation of questions to be addressed, the design of the evaluation study, and the utilization of evaluation results. A paperback, this brief book (160 pp.) offers an excellent introduction to what Weiss refers to as "evaluation research."

49. Windsor, R.A., Baranowski, T., Clark, N., & Cutter, G. (1984). *Evaluation of health promotion and education programs.* Palo Alto, CA: Mayfield.

This text is a useful introduction to the evaluation of health education programs. Windsor et al. have provided readers with a series of health-relevant examples to illustrate their explorations.

50. Worthen, B.R., & Sanders, J.R. (Eds.). (1973). *Educational evaluation: Theory and practice.* Worthington, OH: C.A. Jones.

This volume was one of the earliest compilations of various program evaluation models applied to education. Evaluation theorists whose views are presented in this book include Stake, Cronbach, Scriven, Tyler and others. Worthen and Sanders have authored sections of the book and have included a series of original chapters by a number of evaluation specialists. While focused on educational evaluation in general, the volume is of substantial relevance to program evaluation of health education programs.

51. Worthen, B.R., & Sanders, J.R. (1987). *Educational evaluation: Alternative approaches and practical guidelines.* New York: Longman.

This introductory text is organized around a series of alternative approaches to educational evaluation, including the "objectives-oriented" and "advisory-oriented" approaches.

52. Worthen, B.R., & White, K.R. (1987). *Evaluating educational and social programs: Guidelines for proposal review, onsite evaluation, evaluation contracts, and technical assistance.* Boston: Kluwer-Nijhoff.

This volume provides a first-rate series of practical guidelines dealing with varied aspects of proposal review, onsite evaluation, evaluation contracts, and technical assistance.

53. Zdep, S.M., & Rhodes, I.N. (1977). Making the randomized response technique work. *The Public Opinion Quarterly, 40,* 531-537.

This easily read essay describes the randomized response technique, a procedure used to obtain sensitive information from respondents more accurately than if respondents were directly asked about sensitive information.