EXPERIMENT-RESEARCH METHODOLOGY IN MARKETING

EXPERIMENT-RESEARCH METHODOLOGY IN MARKETING

Types and Applications

Gordon L. Patzer

Experiment-Research Methodology in Marketing

Types and Applications

GORDON L. PATZER

QUORUM BOOKS
Westport, Connecticut • London

Experiment-Research Methodology in Marketing

Types and Applications

GORDON L. PATZER

QUORUM BOOKS
Westport, Connecticut • London

Library of Congress Cataloging-in-Publication Data

Patzer, Gordon L.
Experiment-research methodology in marketing : types and applications / Gordon L. Patzer.
p. cm.
Includes bibliographical references and index.
ISBN 0-89930-960-7 (alk. paper)
1. Marketing research. I. Title.
HF5415.2.P379 1996
658.8'3—dc20 95-24914

British Library Cataloging in Publication Data is available.

Library of Congress Catalog Card Number: 95-24914
ISBN: 0-89930-960-7

First published in 1996

Quorum Books, 88 Post Road West, Westport, CT 06881
An imprint of Greenwood Publishing Group, Inc.

Printed in the United States of America

The paper used in this book complies with the Permanent Paper Standard issued by the National Information Standards Organization (Z39.48-1984).

10 9 8 7 6 5 4 3 2 1

Contents

Preface

Experiments benefit marketing research immensely. A primary reason for this involves the concept of causality—a relationship in which a change in one variable causes a change in another variable (i.e., cause-and-effect relationship). Such relationships cannot be proven with absolute certainty, and often are concluded erroneously. However, unlike other research methodologies, experiments facilitate three criteria required to infer cause-and-effect relationships with reasonable certainty:

evidence of association

appropriate timing

elimination of alternative explanations

This book is divided into three parts. Part I begins with an introductory "real-world" vignette followed by four chapters that present dimensions of experiments. Chapter 1 introduces experiments as a methodology in marketing research and details the concept of causality. Chapter 2 identifies four major components of experiments: hypotheses, subjects, independent variables, and dependent variables. Chapter 3 considers the accuracy of experiments, including experimental error, extraneous variables, demand characteristics, and validity. Since the ability to make valid conclusions is a key aspect of experiments, Chapter 4 examines validity in greater detail.

Many types of experiments are used in marketing research. These types are not always mutually exclusive, and are often combined to design an experiment that is appropriate for the situation. To present these many types of experiments, Part II begins with an introductory "real-world" vignette followed by Chapters 5 through 9. Chapter 5 introduces the topic of experimental design and discusses the settings category of experiments, specifically laboratory and field experiments. Chapter 6 discusses experiments designated as the "independent variable category," which also can be designated the "principle category." Foremost in this category is the basic, classic experiment—the Pretreatment–Posttreatment Control Group Experiment—that is the standard by which other types of experiments are compared. Another experiment in this category is the Factorial Experiment.

Chapter 7 identifies the type of experiment used most frequently in marketing research: the Posttreatment Only Control Group Experiment. It, along with the Solomon Four Group Experiment, focuses on measuring dependent variables. Chapter 8 focuses on subject assignments, specifically the extent of randomization for assigning subjects to experimental treatments. Three experiments in this category are the Completely Randomized Experiment, Randomized Block Experiment, and Latin Square Experiment. Chapter 9 discusses four quasi-experiments that are used frequently in marketing research: One Group Pretreatment–Posttreatment Quasi-Experiment, Static Group Quasi-Experiment, One Shot Quasi-Experiment, and Times Series Quasi-Experiment.

Part III of this book is a culmination and consummation of the material about the dimensions and types of experiments in marketing research. It is culminating because the nine chapters discussed in the prior two parts provide necessary foundation and context for the remaining three chapters: 10, 11, and 12. It is consummating because the material in these three chapters represents the beginning of new territory. That territory is the hands-on application of information discussed in the prior nine chapters.

Chapter 10 provides specific, concrete direction about activities, and stages of activities, necessary to conduct experiments and quasi-experiments. Although the activities are rather infinite, they can be summarized in four stages: (1) formulate the hypotheses, (2) design the experiment, (3) perform the experiment, and (4) present a report. Related discussion about these activities to conduct an experiment includes discussion about the overall marketing research process, including its stages of problem definition and research design.

Chapters 11 and 12 present precise, comprehensive accounts of separate experiments that were conducted to provide information about marketing strategy effectiveness. These accounts go beyond procedural options available to a researcher, to include analysis and discussion of the data for each experiment. The most important difference between these two chapters is first that Chapter 11 demonstrates that an experiment's data can be adequately and usefully analyzed by only calculating and comparing percentages. It then goes beyond percentages by demonstrating the use of analysis of variance (ANOVA), which is a frequently used quantitative technique

performed to analyze data from an experiment. In contrast, Chapter 12 demonstrates a multitude of quantitative tests and techniques which may be used to analyze data from an experiment, and includes additional rationale and explanation about many of the experiment's procedures.

Part I

Dimensions of Experiments in Marketing Research

REAL-WORLD VIGNETTE: BRANIFF AIRLINES

One of the most well-known experiments in marketing research was conducted many years ago by Braniff Airlines. It was used by company management to test a marketing strategy that involved exterior color of airplanes as a means to increase ticket sales. The experiment began with selecting markets with similar characteristics. In some of these markets, the exterior of Braniff airplanes was painted bright, wild colors. In these other selected markets, Braniff maintained its traditional, standard gray-silver exterior color.

Efforts were made so the only difference in the different markets was the exterior color of the airplanes. Therefore, any change in ticket sales that occurred in the markets with brightly colored airplanes could be assumed to be due to the known difference of exterior color. This exterior color difference was maintained for a predetermined period of time. Afterwards, ticket sales were compared for the different markets.

Sales increased significantly during the time of the experiment in the markets with brightly colored exteriors, compared to the other markets. Since all other factors were assumed consistent during this time, company management concluded that the airplanes with brightly colored exteriors caused an increase in ticket sales. This conclusion is reasonable, since it is plausible that passengers were more attracted by the bright colors than the standard gray-silver color. Equipped with this information, the Braniff company painted bright colors on the exteriors of its entire fleet.

How can Braniff know it made the correct conclusion? Was the color of paint the only factor that actually differed between these two sets of markets? Were there other factors outside the researchers' control and awareness that differed between the markets? For example, did the painting of bright colors generate a short-term media awareness that produced increased publicity by local news reports, which in turn generated only a short-term increase in ticket sales?

These are all typical questions related to experiments conducted in marketing research. While marketers can never be absolutely sure that the change in one factor caused the change in another factor, an experiment is the most appropriate research alternative for making such conclusions. In other words, unlike the observation method and survey method, the experiment method does permit, with reasonable certainty, conclusions about the causal relationship between two marketing variables. The ability to make such a conclusion, as well as issues inherent in conducting experiments, regards dimensions of experiments, as discussed in Part I of this book.

Chapter 1

Introduction

The word "experiment" is a commonly used term. Typically, our exposure to formal experiments begins in the context of natural or life sciences and is continued in the news media in the context of medical research. As students we are exposed early to the notion of experiments and most of us have conducted experiments. By the early years of high school we conduct experiments in chemistry classes that measure the effect on one chemical caused by adding another chemical.

The word experiment is heard regularly in the general media in connection with the study of some action or treatment. It is often in the form of news accounts reporting results of medical experiments. An example is a news report of an experiment conducted by the UCLA Medical School to test the effect on reducing heart attack caused by taking an aspirin a day. Another example is an experimental test by the Upjohn Pharmaceutical Company regarding the effect caused on growing hair on a man's bald head by applying Minoxidil (trade name Rogaine).

EXPERIMENTS IN MARKETING RESEARCH

Experiments are conducted frequently in marketing research, as well as in research in medical fields and the social sciences. Regardless of discipline or application, an experiment is a test of a hypothesis that a causal relationship exists between two or more variables. This test focuses on an effect or change that one variable is believed to cause in another variable.

The word experiment is used most frequently in the English language as a noun referring to a study or an investigation of a problem or question. The word experiment is also used as a verb referring to the action of conducting a study or investigation of a problem or question. People who are studied or investigated in an experiment and from whom information is collected are called subjects, in contrast to people in a survey, who are called respondents. A subject is a participant in an experiment, from whom information is collected. Subjects represent some larger group of people to whom information from the subjects is generalized. This larger group of people is likely a target market in which the user of marketing research is interested.

While surveys are the most frequently used method of investigation in marketing research, experiments are also used frequently. Experiments are used in investigations of all the controllable marketing mix variables (e.g., product, price, promotion, and distribution), consumer behavior, and even in regard to marketing research itself. These experiments often incorporate uncontrollable variables of marketing such as the economy and competition, and do so in different applications of marketing such as international marketing, business-to-business marketing, and services marketing.

Test Marketing

Experiments in marketing research are especially prominent in test marketing in order to decide if and how an organization should market a new product. Test marketing is a study that obtains response information through use of an experiment in a realistic marketing setting. While numerous measures of consumer acceptance are conducted (such as product taste, product appearance, package size, price, etc.) before a product is introduced into the actual marketplace, test marketing is the final measure. Experiments for test marketing have proved valuable. One report stated that about 75% of all new products that were test marketed were a success in the actual marketplace, compared to a success rate of only about 20% for new products not test marketed.[1]

Despite the past popularity of using experiments in test marketing, test marketing is the one area in which continued popularity is debated. One survey of 1,200 firms reported that 35% used test marketing in 1975, and 12 years later, in 1987, only 30% used test marketing.[2] The reason is related to the time, money, and less-than-perfect information associated with such applications of experiments.[3] However, these considerations pertain more directly to test marketing than to the topic of using experiments in marketing research.

Future

The future of experiments in marketing research is bright. An article published in 1990 reported that between 1975 and 1987, the number of firms using "formal experimental designs" increased from 16% to 42%.[4]

During this time, firms using "informal experimental designs" increased from 35% to 57%.

There are strong reasons for these increasing statistics. To begin, marketing practitioners constantly encounter questions about cause-and-effect relationships regarding the controllable variables that comprise the marketing mix. These questions often involve specific marketing efforts, concerns about attitudes, and desire for sales. These issues align well with the major components of experimentation, such as hypotheses, independent variables, dependent variables, and conclusions about causality that are not possible with research methods such as observation and survey.

Experiments in marketing research are conducted in regard to every aspect of the marketing mix.[5] Experiments are utilized to investigate product concept, product composition, product appearance, package sizes, colors, prices, distribution outlets, promotion alternatives, advertising appeals, advertising copy, overall effectiveness of different commercials, proper store shelf placement, and on and on. They also are utilized to investigate marketing functions such as professional selling techniques and sales management practices, and even to determine the best number of business-to-business (industrial) sales calls to make.[6] In fact, whenever company management has a question about specific marketing management alternatives, such as one package size versus another, an experiment is likely applicable for investigating the answer. Therefore, it is reasonable to expect their use will increase in all areas, with the possible exception of test marketing. The reason is that company managements are inundated with cause-and-effect questions related to marketing mix variables, when, at the same time, no other research methodology provides as conclusive information about such causal relationships.

TWO KEY FEATURES

Experiments offer two key features not offered by other research methodologies. These are (1) the ability to actually investigate causal relationships and (2) to exert control over major components of the research project. These two features are interdependent such that the ability to investigate causal relationships is due to the extent it is possible to exert research control.

Causality

Causal relationships are the heart of experiments. Regardless of the distinction between people who conduct marketing research and those who use its results, mention of an experiment communicates to everyone the notion of an investigation of the relationship between a particular cause and a particular effect. This notion is in contrast to how potential relationships are investigated through observation and survey research methodologies.

Causality is a relationship in which a change in one variable causes a change or effect in another variable. The first variable is referred to as the independent variable and it causes an effect on the second variable, referred to as the dependent variable. Causality is a topic of special importance in marketing research because marketing management is continually interested in causal relationships. But conducting a marketing research project to investigate the effect on sales ("Y") that is caused by a change in a product or marketing mix attribute ("X") is not nearly as straightforward as may appear initially.

The cause-and-effect relationship of interest to researchers is commonly expressed by the phrase "X causes Y." Cause-and-effect relationship (or causal relationship) is synonymous with causality and, in reality, the phrases or terms are used interchangeably. Likewise, "X causes Y" is a phrase synonymous with causality and, in reality, the two phrases or terms are used interchangeably.

Research theory shows that causality is a complex topic.[7] Understanding the complexity at the level of research theory is not necessary at the current level of marketing research discussed in this book. However, it is important to be knowledgeable about two related fundamentals: (1) the basic expression of causality and (2) the conditions required for causality.

Basic Expression. The basic expression for causality is "X causes Y." Certainty of causation implied at this level is referred to as deterministic causation—a relationship in which "X" always causes "Y." Among the many misuses of the term "causality," implying deterministic causation is the most blatant.

Consider the common misuse of causality, and even deterministic causation, by the news media. The news media routinely (and erroneously) report causality, and even suggest deterministic causation, based merely on the fact that two events occurred during some time period. A specific example is the media's reports that "today's positive (or negative) effects on the stock market were *caused* by" some individual event in the country or world. Sadly, even financial experts voice similar proclamations of cause-and-effect relationships related to the stock market.

Deterministic causation rarely exists. At best, when causation does exist, it is most reasonably to infer probable causation—a relationship in which "X" is likely to cause "Y." When causal relationships are spoken about in marketing (as well as finance, business, social sciences, biological sciences, and life in general), we are most usually referring to likely or probable causation and not deterministic causation.

The difference between probable and deterministic causation is substantial. Therefore, at least three research theory considerations are important to be aware of in regard to the causality expression of "X causes Y":

1. Rarely does only one "X" cause "Y." Instead, it is likely that many variables determine an outcome.
2. The presence of "X" might increase the likelihood of "Y," but increasing the likelihood is not the same as making its occurrence definite.
3. "X" can never be proven with absolute certainty to cause "Y." Based on available evidence it can only be inferred with reasonable certainty.

These three considerations pertain well to marketing and marketing research. First, rarely does only one variable such as package size ("X") cause substantial differences in the level of sales for the product ("Y"). Sales are determined most always by a combination of variables such as the package size, the product itself, its package design, price, promotions, distribution, and so on. This multiplicity of marketing factors equates well with the notion of multiple variables represented by "X_1", "X_2", and "X_3", causing an outcome such as sales represented by "Y."

Second, because a product is marketed in, let's say, the size most desired, which equates with the presence of "X," it might increase the likelihood it will be purchased ("Y"). However, availability of this desired package size does not make its purchase definite, since the marketplace may purchase a competitor's product or may already possess a sufficient quantity of the product.

Third, even though sales of a product may have increased ("Y"), as the product's package size was changed ("X"), marketing executives can never be absolutely certain this change in sales ("Y") was caused by the new package size ("X"). Based on the available evidence, the most that marketing executives can do is infer a conclusion with reasonable certainty, since in the marketing discipline there are many factors that influence product sales. For example, at the same time the product was being sold in a new package size, the competitor's product may have been withdrawn from the market due to a defect, perhaps a special promotion campaign was begun by the competitor, possibly the price was lowered, or maybe the economy changed, and so forth.

Required Conditions. Research theory shows causality cannot be proven or concluded with absolute certainty. It does show, however, that causality can be inferred and subsequent conclusions made with reasonable certainty. According to research theory,[8] evidence of association, appropriate timing, and elimination of alternative explanations are all required to make even reasonable conclusions about causality.

Evidence of Association. For legitimate conclusions that causality exists, two entities or events must be associated in a manner referred to as concomitant variation. Concomitant variation is the extent to which a cause ("X") and an effect ("Y") occur together or vary together.

While concomitant variation is a required condition to infer causality, it alone is not a sufficient condition. For example, the fact that a change in the container size of Heinz ketchup occurred at the same time as the product's sales increased is concomitant variation. They both occurred or varied together. However, other events in the marketplace could have occurred at the same time as the package size change, and these other variables could have caused or contributed to the change in sales.

Appropriate Timing. Appropriate timing is required to make legitimate conclusions that causality exists. Specifically, a cause ("X") must occur before or simultaneously with the occurrence of an effect ("Y"). The required order is that a cause cannot occur after an effect. Consider the situation in which sales increase during the month of June for Heinz ketchup, and afterwards, in September, a new container size of the product is distributed in the marketplace. Obviously, it is not logical to conclude that this increase in product sales was caused by the change in package size.

Sometimes the timing of events is less easily identified. In these situations care must be exercised to determine the order in which the events of interest occur. For example, maybe customers became aware that a new, less desirable size of container would replace the current desirable size container. In response, the purchasers rushed out to purchase the current size, with the effect of increased sales actually being caused by current anticipation of a future event.

Lack of Alternative Explanations. Alternative explanations for what may cause an effect ("Y") must be eliminated if legitimate conclusions are to be made about the existence of causality. If more than one explanation can explain why an event occurred, then a conclusion of causality cannot be made about a specific cause.

Consider that during the month of March a new container size of Heinz ketchup is marketed. Also in March a new advertising campaign is begun on television and a new point-of-purchase display is set up in the grocery stores. Then, in April, the Heinz company experiences a 10% increase in its ketchup sales. It cannot be concluded that the new package size caused the increase in sales (i.e., the effect) because the advertising campaign and the point-of-purchase display also were new during that time and, therefore, also offer explanations for the increase in sales. In order for a legitimate conclusion to be made about the causality between the package size change and the product sales, it is necessary that all other factors but one remain unchanged during the study of the relationship between these two variables of interest.

Being able to eliminate alternative explanations in relationships believed to be causal is of extreme importance to marketing research in particular and business in general. For an example of the relevance in the advertising and broadcasting industry,[9] see Exhibits 1.1 and 1.2, which demonstrate how this importance translates into ongoing issues with immediate financial

Exhibit 1.1
The Importance of Eliminating Alternative Explanations: A Firsthand Experience Example

The numbers of people who watch television programs concern many people. They range from small-town business people who buy advertising time on their local television station (network affiliate) to executives at major corporations who buy advertising time on the four television networks (ABC, NBC, CBS, and FOX).

There are at least three reasons for these concerns. First, television networks and location television stations only broadcast programs which attract large numbers of viewers. Second, advertisers pay the networks and stations based on the number of viewers for each program. Determining the number of viewers for any one program is consequently an important marketing research business that is complicated, given the approximately 260 million people spread across the United States.

The most widely used service for measuring audiences of American network television is the A.C. Nielsen Company. In 1990, each Nielsen ratings point represented 921,000 homes. Since these ratings influence huge payments that advertisers make to the television networks and ultimately to the local television station, the Nielsen rating service is always trying to get the most accurate measure possible. After several years of declining viewership attributed to a growing number of cable television stations and VCRs, the Nielsen ratings showed a substantial change at the end of one year. Much to the

Exhibit 1.1 (continued)

pleasure of ABC, CBS, and NBC, Nielsen television ratings for that year showed a substantial increase in the number of network viewers.

Advertisers demanded to know the cause ("X") of the increased Nielsen ratings ("Y"). Were the increased ratings caused by an actual increase in the number of viewers, or were other explanations of the cause possible? The task for conclusions about this cause-and-effect relationship was to eliminate alternative explanations. These conclusions meant huge differences in the amount to be paid for advertising time.

In response to the advertisers' demand to know the cause, the explanation offered by the television networks and stations was that the increased ratings were caused by an actual permanent change in television viewing. This change was in turn caused by (1) better programs that generated greater interest among viewers to watch more television, and (2) the increased proportion of working wives, which meant more family time was being spent at home watching television.

However, a separate marketing research project during this same time indicated growing dissatisfaction with television programming. Therefore, advertisers offered other possible alternative explanations for the cause of the increased ratings. They suggested a statistical aberration due to a change at the same time in the

Exhibit 1.1 (continued)

sampling-measurement procedures by the Nielsen Company, a temporary change caused by one-time-only special event programming by the networks, or a transitional change caused by external factors such as the weather and economy. A lengthy investigation was conducted.

The conclusion stated the most likely explanation of the cause for the increased ratings was a temporary increase in viewers due to an extended period of inclement weather in the nation.

consequences. This advertising example and the above Heinz ketchup example are problems of trying to eliminate alternative explanations after the fact. These two examples illustrate the value of being able to eliminate alternative explanations before the fact. Experiments make it possible to accomplish this task because of a major feature known as "research control."

Research Control

A hallmark feature of experiments is the ability to make conclusions about causality with reasonable certainty. The aspect of experiments that makes these conclusions possible is the researcher's ability to exert greater research control than is possible with other research alternatives. Research control is the ability to manipulate research procedures in a manner that eliminates alternative explanations for the findings. These manipulations involve the variables believed to cause the effect in which company management is interested.

Company management at Procter & Gamble may believe that the variable of package color has a significant effect on sales of toothpaste. It may wonder if changing the current, primarily white container of Crest toothpaste to a subtle green color would cause more product sales. In this example, the consequent research problem, no doubt defined earlier from the product definition stage of the marketing research process, is to investigate if a green container produces a higher level of sales. Research control is an important feature of experiments, regardless of whether the results indicate support or rejection of the belief. Therefore, if the results in this example

Exhibit 1.2
The Continuing Importance of Eliminating Alternative Explanations for Causal Relationships

AND QUESTIONING ABOUT CAUSE-AND-EFFECT CONTINUES

In the prior example, alternative explanations for an *increase* in the ratings of television viewers were offered and explained. Five years later the situation had reversed itself. Alternative explanations were now being offered and explained for a *decrease* in the television ratings.

An example of this decrease is expressed in an Associated Press headline that read: "Lowest [television viewing] ratings in history..." That news article reported that 10 years earlier, before the advent of cable television, the three major television networks and their associated local television stations accounted for 90% of the nation's television viewers, but during the just completed year this viewership had dropped another 4% to only 65% of the viewers. The article then reported: "But [television] officials differed Tuesday on whether cable and Fox or Nielsen's methods of estimating viewers caused the drop."

indicate that a green toothpaste container does increase sales, the researcher could infer with reasonable certainty that a respective conclusion is accurate.

Unique Feature. Research control is a feature unique to experiments. It is not even part of the methodology for observation or for surveys. Consider the above example in which the company management at Procter &

Gamble believes a different color toothpaste container will cause a higher level of product sales for Crest toothpaste. An observation research project would limit researchers to observing consumers purchasing and not purchasing the current product (with its white package). Based on these observations, it would be impossible to make accurate conclusions about purchases of the product if it was in a green container.

A survey research project could provide more information than an observation, but it still would be less information than possible with an experiment. A survey would permit researchers to ask questions of the consumers who now purchase and those who do not now purchase Crest toothpaste. These questions could include asking why they do or do not purchase the current product and if they would purchase the product if it was marketed in a green container. However, based on these responses, it likely would be next to impossible to make accurate conclusions about purchases of the product in a new, green container.

The problem is that consumers really do not know what they would do in this case. Even if they did, very few consumers in this situation would likely admit to changing their purchases of a product in which only the package color was changed; and especially from its current appealing white color to a novel new green color. But yet, such changes as package color represent marketing mix variables that are frequently manipulated in attempts to cause increased sales of products.

Manipulations. Part of research control is the ability to manipulate the topic or variable of interest. In an experiment such as the above example of Crest toothpaste, green containers could be actually placed on retail store shelves. Response to the green container in terms of actual sales then could be assessed for the duration of the experiment. From these measures, more accurate inferences about actual marketplace behaviors could be made than are possible with research investigations relying on observations or surveys.

This more accurate information is due in part to the control which permits the variable of interest to be disguised. An experiment allows the researcher to conduct an investigation without the participants being aware of the purpose. This disguise incorporates the benefit of observation research that is unobtrusive in the subject's life and the benefit of survey research to focus on the variable of interest, without the negatives of having to rely on indirect behavior as in observation, or possible false statements as in a survey. Participants are then unable to, intentionally or unintentionally, give an answer they feel is most logical, least embarrassing, or the answer that is expected. This feature of disguise is especially important for marketing research because we know that consumers, individually and collectively, frequently do not respond to marketing efforts in a logical or socially expected manner.

NOTES

1. "Test Marketing: What's in Store," *Sales and Marketing Management* 128 (March 15, 1982), pp. 57–85.
2. Danny N. Belleger, Roy D. Howell, Jim Wilcox, and Barnett A. Greenberg, "The Diffusion of Marketing Research Techniques in Business: Implications for Marketing Education," *Journal of Marketing Education* (Spring 1990), pp. 24–29.
3. Test marketing represents only one application of one type of an experiment. As such, a thorough discussion of test marketing is outside the scope of this book regarding the experimental methodology in marketing research. For the pros and cons of test marketing, the reader is referred to more basic books, and university textbooks, dealing precisely with test marketing within such topics as new product development and marketing management.
4. Belleger et al., "The Diffusion of Marketing Research Techniques in Business," pp. 24–29.
5. For a thorough discussion of the use of experiments in marketing see David M. Gardner and Russell W. Belk, *A Basic Bibliography on Experimental Design in Marketing* (Chicago: American Marketing Association, 1980).
6. Ibid.
7. For a more generic and thorough discussion of causality see Claire Selltiz, Lawrence S. Wrightsman, and Stuart W. Cook, *Research Methods in Social Relations*, rev. ed. (New York: Holt, Rinehart and Winston, 1959), pp. 80–82; David A. Kenny, *Correlation and Causality* (New York: John Wiley, 1979).
8. See Selltiz et al., *Research Methods in Social Relations*, pp. 80–82.
9. This example is based in part on the following two articles by the Associated Press (New York): "Lowest Ratings in History for Three Networks," *Des Moines Register*, April 18, 1990, p. 5A; and Kevin Burns, "TV Viewing Up—Maybe," *Media Message: An Ogilvy & Mather Commentary on Media Issues* (October 1983), pp. 1–7; and one broadcast news report by Grant Perry in New York, "The Networks Say the Problem Is with the Messenger," *CNN Business News*, Atlanta, GA, May 4, 1990.

Chapter 2

Major Components of Experiments

An experiment is comprised of four major components: hypotheses, independent variables, dependent variables, and subjects. Regardless of discipline, context, or situation in which an experiment is conducted, these four components comprise the core of an experiment. Their precise designation is in response to the outcome of the problem definition stage of the marketing research process. If that outcome suggests conducting an experiment, it is typically because of rather precise statements (i.e., hypotheses) that reflect relationships believed to exist between two variables among a population of interest to company management.

The hypothesis serves as a guide for the other components and, therefore, represents the first decision by the researcher about these four components. It specifies the relationship of interest between two variables, which are the independent variables and dependent variables among a certain population. While the relationship pertains to the population, conclusions about the relationship and population are based on a subset or sample of that population known as subjects.

HYPOTHESES

A hypothesis is an unproven proposition. The proposition is usually a statement about reality that is empirically testable. It should be in direct response to the question(s) formulated in regard to the problem definition stage of the marketing research process.

Hypotheses in marketing research commonly regard a specific type of question that focuses on the causal relationship thought to exist between two or more variables. An example of these relationships is those which may exist between an effect on marketplace behavior (such as attitude, intent to purchase, or actual sales of a product) that is caused by an aspect of the marketing mix (such as a product's composition, its price, advertising program, and/or distribution).

When the problem definition stage focuses on causal relationships, an experiment with its ability to test specific hypotheses is commonly the most appropriate choice of research method. This choice is because the research control associated with an experiment allows the greatest certainty when making conclusions that reject or support a hypothesis about marketing dynamics. For example, the problem definition stage may yield questions about the effect on marketing effectiveness that is caused by different product designs, prices, distribution outlets, or promotion approaches.

There are literally thousands of potential marketing strategy and corresponding marketing research questions that can be addressed through experiments via hypotheses. A research question to be answered may, for example, focus on the impact of package color on marketing effectiveness. Since the researcher can control the color of the package, the hypothesis in a respective experiment would focus on: "Does package color cause an effect on product sales." Thus, an experiment for Procter & Gamble could test the hypothesis that: "When given a choice, more consumers will purchase Crest toothpaste in a green container than a white container."

INDEPENDENT VARIABLES

An independent variable is a factor which the researcher controls. It is called independent because the researcher can choose what it should be and can manipulate it to study the effect it causes. The independent variable is specified in the hypothesis. For example, since hypotheses in marketing research commonly relate to aspects of the marketing mix, the independent variable is frequently a controllable variable of the marketing mix, involving the product, price, promotion, and/or distribution. In the above hypothesis about Crest toothpaste, the color of the toothpaste container is the independent variable which can be manipulated or controlled by the researcher to be white, green, red, or any other color of interest to company management.

Multiple Values

The independent variable is commonly comprised of more than one value. An independent variable value is the specific manipulation being studied in regard to the independent variable. In the Crest toothpaste example, the container color is the independent variable and the specific colors tested (white and green) are the independent variable values. Other

independent variable values (i.e., other colors) could be added to this Crest toothpaste experiment.

A strength of experiments is that different values of an independent variable can be tested at the same time. As a consequent, the effects of each tested value can be determined. To do so, different participants in an experiment (i.e., subjects) are exposed to different values of the independent variable known as experimental treatments.

An experimental treatment is the independent variable value to which subjects in an experiment are exposed. Experiments with more than one independent variable value are said to have the respective number of experimental treatments. At times, the term experimental condition is used in place of experimental treatment. Therefore, for practical purposes, experimental condition is synonymous with experimental treatment and, in reality, these two terms are used interchangeably.

An independent variable can have many values. Management may question which of three values or four values is most effective. If the independent variable is package color, four values could be green, red, yellow, and blue. The number of values that an independent variable has in an experiment is theoretically unlimited. But the difficulty in conducting an experiment increases as the number of values for an independent variable increases. This difficulty is due to an increased number of participants (subjects) required to conduct the experiment, increased expenses and problems to collect the data, and increased efforts for analyzing and interpreting the results. Consequently, for practical purposes, experiments typically are conducted with a maximum of two to four values for an independent variable.

Exhibit 2.1 illustrates an experiment with one independent variable (package color) with three values (red, green, and yellow). This figure shows that purchase values of the product under study are tabulated separately for the red packages, green packages, and yellow packages. It also shows that in an experimental design such as this one, the data analysis permits comparison of purchase data for each of the different independent variable values.

Quantitative and Qualitative Values

An independent variable can have a single value or multiple values that are qualitative as with labels, or quantitative as with numerical amounts. Qualitative values and quantitative values are both involved in experiments conducted in marketing research. A marketing research experiment involving qualitative values for an independent variable could be advertisement appeal (humorous and serious), retail store (specialty store and mass discounter), or package color. In the Crest toothpaste example, the white and green container colors represent two qualitative values of the independent variable, and together they represent a categorical variable—an indepen-

Exhibit 2.1
Prototype Table Illustrating an Experiment with One Independent Variable That Has Multiple Values

Table # [actual number to be provided by researcher]

Intent to Purchase Product by Package Color
[actual title to be provided by researcher]

COLOR		
Red	Green	Yellow
*(#1)	*(#2)	*(#3)

*The space where the asterisk is located represents a different cell for each of the three values for which respective data are tabulated, analyzed, and reported.

dent variable with a limited number of possible values. Related classification variable is synonymous with categorical variable and, in reality, the two terms are used interchangeably.

Independent variable values in terms of quantitative amounts could be package size (e.g., 2 ounces and 8 ounces), number of salespeople to employ to cover a particular territory (e.g., 10 and 25), or the amount of advertising budget to spend on print media during a certain period of time. In an experiment for Crest toothpaste, the quantitative values of an independent variable involving advertising budgets could be, for example, $200,000 and $500,000. Such a variable represents a continuous variable—an independent variable with an unlimited number of possible values.

Prototype Tables

Exhibit 2.1 illustrates an experiment with multiple independent variable values. This figure serves a second purpose in that it also illustrates a valuable research tool referred to as a prototype table (or dummy table). Prototype table is a visual aide for understanding what data are to be collected and how these data then will be tabulated, analyzed, and presented. Since several stages of all research projects can benefit by a prototype table (e.g., data collection, data analysis, and report presentation), constructing prototype tables is an important activity that should be conducted thoroughly early in a marketing research project, such as at the research design stage.

The prototype tables should be complete with the variables and measures that will comprise the table, the table number, and the table's precise title and headings. The only items that should remain to be filled in after the research design stage are the actual statistics for the data that are collected.

The prototype table in Exhibit 2.1 uses asterisks to designate cells. A cell is a location in a table or chart in which data are entered. As the table title and column headings indicate, data regarding intent to purchase the product in a red package would be entered in cell #1, for the green package in cell #2, and for the yellow package in cell #3. Similarly, actual purchase data for a product might be entered. Consider an experiment conducted over a one-month time period for Crest toothpaste. Measuring the sales in one neighborhood that stocked three different container colors in each store showed that consumers purchased 1200 tubes of Crest toothpaste in red containers, 900 in green containers, and 1500 in yellow containers. These purchase quantities could then be entered in the table accordingly, which would provide a general visual understanding of how the data (sales and package color) relate to each other.

Prototype tables are not limited to sales data. Measures differ according to research objectives, and all the many measures from experiments that differ according to research objectives can be placed into prototype tables. Instead of sales of Crest toothpaste, an experiment could measure consumer attitude, product perception, awareness, and/or intent to purchase (see Exhibit 2.2), and so on. As this exhibit indicates, after data from all 150 customers are collected, one mean value representing each of the three respective groups of customers would be entered in the respective cells. If the mean attitude (on a 10-point scale) is 8.2 for the red container, 7.3 for the green container, and 9.0 for the yellow container, these data values then would be entered into the table accordingly.

It should be noted that this table in Exhibit 2.1, with only one independent variable, as represented by the columns, is not as common as tables with more than one variable. For example, Exhibit 2.4 is a prototype table illustrating more than one independent variable and each with more than one value. In this latter prototype table, a cell still represents a location in which data values are entered, but now that location is a combination of two or more variables, such as package color, product price, and subject gender.

Experimental Group and Control Group

Company management often becomes interested in distributing new versions of a product, and will turn to marketing research with related questions. An experiment permits management, with the assistance of the researchers, to determine the effectiveness or impact of each of these versions in their attempt to identify the best to market. But before making a

Exhibit 2.2
Prototype Table for Experiment Reporting Data Other Than Sales Figures

```
Data for a prototype table need not be actual sales data.
An alternative experiment to investigate the best
container color for CREST toothpaste could occur earlier.
This experiment could involve a new CREST container mock-
up displayed to consumers/subjects in a laboratory, rather
than actual new containers stocked on store shelves.
Shoppers in a local mall could be invited to participate
in a study. If they agree, they are escorted to a room
where one group of 50 customers are shown a tube of CREST
toothpaste in a red container mock-up. In the same manner,
a group of 50 different customers are shown a green
container mock-up, and another group of 50 customers are
shown a yellow container mock-up.

After each exposure to a container color mock-up, attitude
and purchase intent measures are assessed. When data from
all the selected customers (i.e., subjects) are collected,
one mean value representing each of the three respective
groups of customers could be entered in the respective
cells to provide a visual aid to understanding the general
effect caused by the container color. The resulting
prototype table might appear as follows:
```

Table 1. Mean Values for Purchase Intent, According to Container Color for CREST Toothpaste.

CONTAINER COLOR		
Red	Green	Yellow
8.2	7.3	9.0

Note: 10-point scale was used, 1 representing no intent and 10 representing definite intent.

change in their existing product, the company should know if the proposed changes are likely to have a positive or negative impact on consumers, and to what extent, compared to their current product version.

Changing from a successful product version may cause detrimental effects if the new version is less well received by the consumer. As a result, marketing researchers should always be interested to determine the effects of potential changes in a marketing mix compared with the status quo. To permit this comparison, two types of subjects exist in most experiments: the experimental group and the control group. An experimental group is the group of subjects who receive an experimental treatment. A control

Exhibit 2.3
Prototype Table Illustrating an Experiment with One Independent Variable That Has Multiple Values and a Control Group

Table # [actual number to be provided by researcher]

Intent to Purchase Product by Package Color
[actual title to be provided by researcher]

COLOR			Control Group
Red	Green	Yellow	
cell #1	cell #2	cell #3	cell #4

group is the group of subjects who do *not* receive an experimental treatment. The purpose of the control group is to serve as a standard against which possible new marketing actions are compared with the actions currently being performed, or not performed. Therefore, during an experiment, subjects in the control group are not exposed to any of the independent variable values and are treated in every way possible the same as in the past, which is to maintain the status quo for these subjects.

The Crest toothpaste example (portrayed in Exhibit 2.1) illustrated three experimental treatments that were tested: red container, green container, and yellow container. Crest toothpaste was already successfully marketed for many years in a primarily white container. In this situation, management at Procter & Gamble would definitely question the impact of the new proposed colors (e.g., perception of the product) relative to Crest toothpaste in the traditional white container. The respective experiment might involve four different cities or markets. In three of the markets, Crest toothpaste would be sold in one of the three new container colors (red, green, or yellow), but in the fourth market the product would continue to be sold in its traditional white container. The three markets with the new container colors (i.e., those subjects receiving an experimental treatment) are the experimental groups. The one market with the traditional white container (i.e., those *not* receiving an experimental treatment) is the control group.

Exhibit 2.3 illustrates this current experiment where, in addition to three experimental groups, there is also a control group. Cells #1, #2, and #3 represent experimental groups, since subjects in these cells receive an ex-

perimental treatment of a new container color: red, green, or yellow. Ce
#4 represents the control group in this experiment, since subjects in thi
cell do not receive an experimental treatment. During the experiment, th
fourth group is treated the same as in the past, which is to provide th
package in the traditional white color. Purchases of the product during th
time period of the experiment are recorded for this control group, as we
as the experimental groups. All these data are then compared in the dat
analysis stage to determine which experimental treatment, or more precise
which experimental condition, causes the greatest sales.

Multiple Independent Variables

The independent variable in an experiment can have more than or
value. An experiment also can have more than one independent variab
and, in fact, is comprised commonly of more than one independent vari
ble. A reason for multiple independent variables is that experiments allo
researchers the ability to test the effect caused by multiple independe
variables, individually and collectively. For example, in addition to an i
dependent variable such as package color, company management al
might be interested in such independent variables as product price ar
advertising message. Another example could be company management
Procter & Gamble who may have questions about which of three variabl
or four variables is most effective. These variables could be container colo
product price, package shape, and version of advertisement.

While the number of independent variables in an experiment is theor
ically unlimited, the difficulty of conducting an experiment increases as t
number of variables increase. The reason is that more participants (subjec
are required, which translates into increased difficulty in collecting a
analyzing the data and interpreting the results. Therefore, for practical pu
poses, experiments are typically conducted with a maximum of about fo
independent variables.

Multiple Variables and Multiple Values

An experiment can be conducted to investigate independent variables a
accompanying values at the same time. In other words, both multiple i
dependent variables and multiple values for each variable can be inves
gated in one experiment. A strength of an experiment is that it can ident
differences in effects caused by different independent variables and differ
independent variable values. An even greater strength is the ability to id
tify differences between the independent variables while at the same ti
identifying differences between each value of each variable.

Variables to be investigated stem from research questions to be answer
Related, these questions that are delineated in the problem definition st

Exhibit 2.4
Prototype Table Illustrating an Experiment with Multiple Independent Variables and Multiple Values for Each

Table # [actual number to be provided by researcher]

Intent to Purchase Product by Package Color by Product Price by Consumer Sex
[actual title to be provided by researcher]

	Color								
	Red			Green			Blue		
	Price			Price			Price		
Sex	H	M	L	H	M	L	H	M	L
MALE	*	*	*	*	*	*	*	*	*
FEMALE	*	*	*	*	*	*	*	*	*

*The space where the asterisk is located represents a cell.

of the marketing research process are practically limitless. As a consequent, a large number of hypothesized relationships can stem from questions pertaining to, say, only three independent variables such as package color, product price, and commercial length. Company management might question which of three package colors is best (e.g., red, green, yellow, and so forth), which product price is best (e.g., $2.99, $3.49, $3.99, etc.), and which commercial length is best (e.g., a 60-second commercial, 30-second commercial, 15-second commercial, and so on).

Exhibit 2.4 illustrates an experiment with three different independent variables (package color, product price, and consumer sex). This prototype table is similar to that presented in Exhibit 2.1. The difference is that more cells are shown because of the more independent variables and more independent variable values. It should be noted that 150 subjects are necessary in regard to the prototype table in Exhibit 2.1, with 3 cells times 50 different subjects comprising each cell. But in an experiment such as the one represented by the prototype table in Exhibit 2.4, 900 subjects are necessary if 50 different subjects comprise each of the 18 cells.

DEPENDENT VARIABLES

Up to this point, causality has been discussed in terms of cause and effect. The independent variable regards the cause, and it is hypothesized to caus an effect which regards the dependent variable. A dependent variable is a variable that is responsive to (i.e., depends on) an independent variable. Th This response is the effect that is caused by an independent variable. Th prior prototype tables (Exhibit 2.1, Exhibit 2.3, and Exhibit 2.4) illustrat the relationship between independent and dependent variables. In these fig ures, the cells are the locations where the dependent variable values ar recorded. A dependent variable value is a numerical measure used for pur poses of analysis relative to an independent variable. Two common de pendent variable values in marketing research are sales volume figures an attitude scale numbers assessed before and/or after participants in an ex periment are exposed to an experimental treatment.

The dependent variable in marketing research experiments commonl pertains to response aspects of consumer behavior and market behavio The reason is that the dependent variable typically involves the respons portion of research questions posed in the problem definition stage. In othe words, questions about consequences or effects in response to various mar keting variables that arise from the problem definition stage.

Furthermore, when the problem definition stage focuses on aspects o marketplace responses (such as attitudes, purchase decisions, and sales vo umes), the logical choice of a research methodology is an experiment. De pendent variables are then inherent in this methodology. In fact, th importance of dependent variables is heightened, since no other researc methodology permits the same level of certainty for making conclusion about hypothesized effects that an independent variable has on a depender variable.

SUBJECTS

Marketing research experiments hypothesize relationships between in dependent and dependent variables. Integral to the dependent variable ar the entities on which the hypothesized effect occurs and is measured. Thes entities are referred to as subjects. Formally defined, a subject is a partic pant in an experiment. Like all marketing research, the subjects in an ex periment are a sample or a subset of the population of interest, which typically a company's target market about which a marketing decision mu be made.

Subjects in marketing research experiments are usually people, but sub jects also can be nonhuman entities and inanimate entities. Nonhuman en tities are animals such as dogs or cats that could be subjects in a marketin research experiment when, for example, the research focus is on the r

sponse of pets rather than their owners. Inanimate entities used in marketing research are items such as stores and sales territories, in which case they are commonly referred to as test units. A test unit is the term usually reserved to describe nonhuman, inanimate participants in a research project. However, researchers on occasion also use the term "test unit" to describe human subjects. In reality, the two terms—subject and test unit—are synonymous and can be used interchangeably when it is grammatically appropriate. Regardless of descriptive label, measures of dependent variables are collected, recorded, and analyzed from both subjects and test units.

Treatment Groups and Control Groups

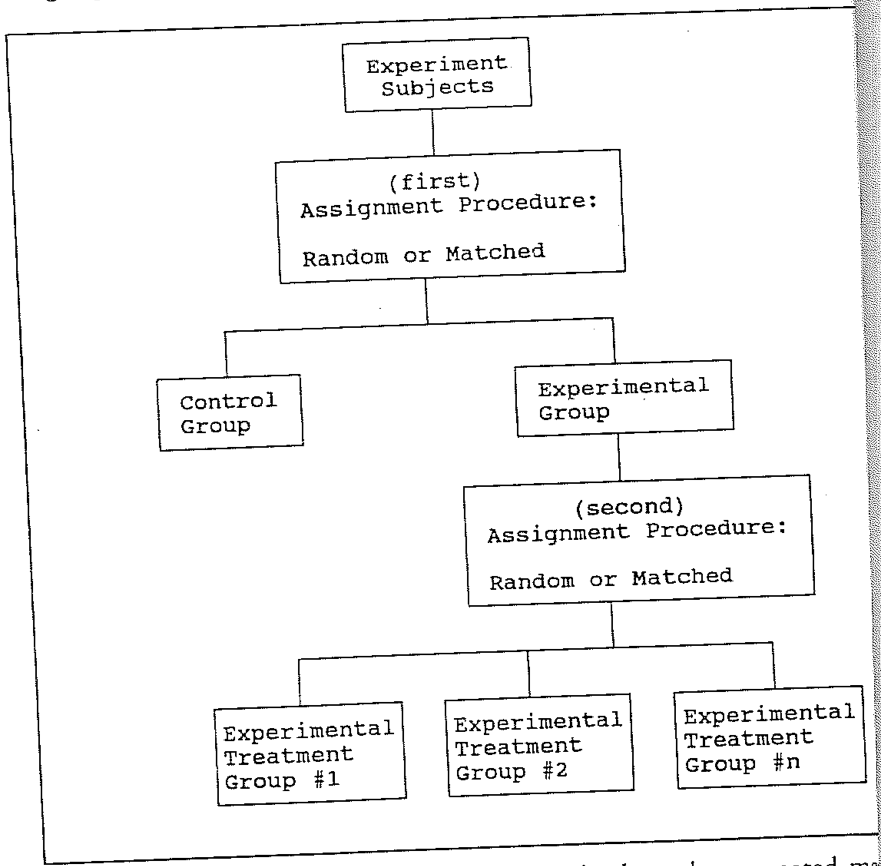
Two broad categories of subjects exist in an experiment: (1) those in the treatment group and (2) those in the control group. Treatment group subjects are those subjects exposed to one or more of the experimental treatments. Generally, there is more than one treatment group of subjects, since a different group is assigned to each of the different independent variable values. Control group subjects are those subjects who are not exposed to any of the experimental treatments and are included in the experiment for comparison purposes. Generally, there is one control group of subjects who serve an essential experimental function of providing baseline measures.

A baseline measure is the results of measures under normal conditions. In an experiment, measures of a dependent variable for a control group of subjects (i.e., those who receive no experimental treatment) are the values against which measures of the dependent variable for a treatment group are compared, to determine the treatment's effect. Because baseline measures by definition refer to normal conditions, as opposed to experimental conditions, these measures are often abbreviated and simply referred to as the norm or norms. Norm is synonymous with baseline measure and, in reality, these two terms are used interchangeably.

Assignment Procedures

Subjects selected for an experiment often are divided two times (see Exhibit 2.5). The first time subjects are divided is to assign them, through either a random or matched procedure, to the control group or experimental group. Assuming there is not more than one experimental condition, the second level division is to then assign subjects to particular experimental treatments (or experimental conditions). At both times, subject assignment is a systematic division of subjects among experimental groups. A systematic division is used because subjects that comprise the sample of an experiment are not treated the same. It is therefore necessary to determine which subjects receive which treatments. To make this determination, the total sample is divided through a process of either a random assignment or

Exhibit 2.5
Assigning Subjects to Experimental Conditions



matching assignment, and much less frequently through a repeated me sures process. When the experiment has more than one experimental grou those subjects assigned initially to the experimental group are divided second time, again through either a random or matched procedure, to o of the experimental treatment groups.

Random Assignment

Random assignment divides subjects among experimental treatme based on a random process. Formally defined, random process is a proc dure to select objects, in which all objects have an equal chance to selected. There is no pattern in which individual objects are selected. common two-step procedure to perform a random process is (1) assign number to each object, person, entity, and so on, and then (2) go to a

location in a random numbers table (see "Table of Random Numbers" in appendix) to determine the order of selecting the objects for the desired purpose.

Through this "control by chance,"[1] certain types of subjects are distributed equally among the experimental treatments so that subjects comprising each of the different experimental conditions are the same. Therefore, if different experimental conditions yield different results, the results can be attributed to differences of effects caused by experimental treatments, rather than differences in subjects who comprised the different groups.

Consider an experiment for Bounty paper towels. The independent variable is towel design, experimental treatments are "fancy" and "plain," and the subjects are 100 housewives from Columbus, Ohio. The random assignment procedure first identifies each of these subjects with a number. Then, using a random digit table, individual subjects are assigned to the "fancy" package condition and to the "plain" package condition.

Generally, less-than-perfect information is better than no information. But consider the alternative assignment procedure described in Exhibit 2.6 for Bounty paper towels. In this situation, less-than-perfect information is worse than no information, since it likely will mislead the decision maker.

Matching Assignment

In some situations, a researcher makes an informed decision not to use random assignment. This alternative to random assignment is matching assignment. Matching assignment divides subjects among experimental groups based on an identifiable factor, believed to produce a consistent effect in response to the independent variable. Rather than assign a subject to an experimental treatment based entirely on equal probability, subjects are assigned based on a specific factor or characteristic. This factor is a characteristic that can be identified by the experimenter and is one which the experimenter believes has a consistent effect in response to the independent variable.

Consider again the Bounty paper towels experiment. The experimenter may know that women working outside the home strongly prefer "plain" designs rather than "fancy" designs. It is then best to control their assignment, to ensure that an equal number of these subjects are assigned to the "fancy" condition and to the "plain" condition. This control does not determine which individual subject or group of subjects is assigned to a specific condition. It only determines that an equal number of this type of woman is assigned to the different conditions.

In the Bounty paper towel experiment, subjects would first be identified as belonging to one of two groups: Either working outside the home or not working outside the home. Then within these groups, an equal number are assigned by a random process to either the fancy condition or the plain

Exhibit 2.6
Assignment by Arrival Time: An Alternative to Random Assignment

An alternative to random assignment for the Bounty paper towel experiment is to assign subjects according to their arrival time at the testing center. The basic dimensions of the experiment would be the same, with towel design as the independent variable, "fancy" and "plain" as the experimental treatments, and 100 housewives from Columbus, Ohio, as the subjects.

This alternative procedure might assign the first 50 ladies to arrive to the "fancy" package condition and the next 50 to the "plain" package condition. However, it is reasonable to speculate that consistent differences could exist between these two groups. The first group may be predominantly women who do not work outside the home and their more flexible schedules permit them to arrive early. The second group may be predominantly women who work outside the home and their less flexible schedules restrict their arrival time. Possible implications for the Bounty paper towel experiment is that the first group might be more concerned about fancy-type paper towels that "dress up" the kitchen. In contrast, the second group might be more concerned about paper towels that serve the function of drying than its appearance.

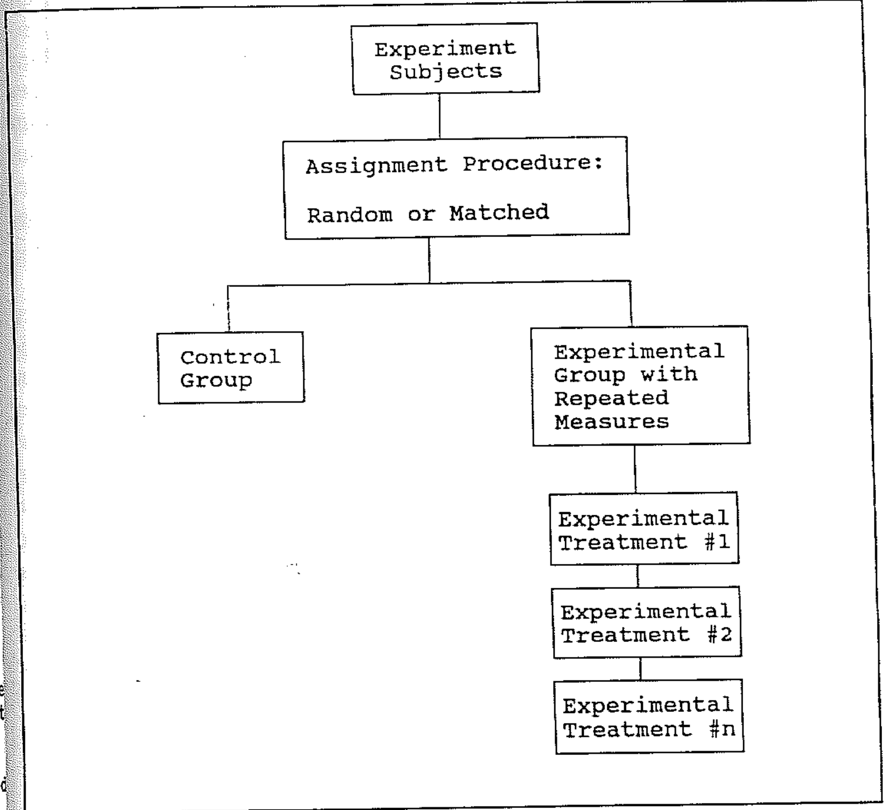
The consequent is that if the two different experimental treatment conditions yield different test results, the researcher could not make valid conclusions. It would be unclear if the data differences were due to the type of paper towels and/or to the type of subject assigned to the respective conditions. If the researcher did conclude that one type of tested paper towel was better received than the other, and likely would generate greater sales, subsequent marketing decisions by company management could be erroneous.

condition. Random assignment is therefore not omitted from the matchin assignment process. It is performed, rather, as a second stage of subje assignment.

Repeated Measures Assignment

A third procedure for subject assignment, which is used on a limite basis, is repeated measures assignment. Repeated measures assignment a signs subjects to more than one treatment condition in the same expe ment. Rather than divide subjects among the different experiment treatments, as with the random and matching assignment procedures, t same subjects are assigned to different experimental treatments. As Exhib 2.7 shows, subjects are assigned to the control group or experimental gro

Exhibit 2.7
Assigning Subjects to Experimental Conditions with Repeated Measures



through either a random or matched procedure. Once assigned to the experimental group, these subjects are exposed to all, or at least more than one, of the experimental treatments.

There are pros and cons for using a repeated measures assignment procedure. One advantage is that particular treatments are not disproportionate or inappropriately comprised of a certain type of subject. In this way, repeated measures achieve the goals of both random assignment (for proportionate composition of subject types) and matching assignment (for appropriate composition of subject types).

A second advantage is that fewer subjects are necessary to conduct an experiment. To obtain 50 responses to the "fancy" package design and 50 responses to the "plain" package design in the above Bounty paper towel experiment, 100 subjects are needed. But if repeated measures are assessed, 50 responses to each experimental condition can be obtained by utilizing the same 50 subjects for both treatments.

The major disadvantage of repeated measures is that, by definition, subjects are exposed to more than one experimental treatment. This multiple exposure defeats a major strength of experiments. Specifically, research control that permits the research purpose to be disguised is very likely to be lost. Integral to an experiment is the fact that only the independent variable changes between experimental treatments. With repeated measures assignment, subjects are first shown one experimental condition and then the second. The independent variable (i.e., the research purpose) then becomes apparent to the subjects, and their responses may be affected accordingly. Ultimately, the loss of this research control, which is an experiment's strength, severely reduces the ability to make conclusions about cause-and-effect relationships.

The disadvantage of repeated measures translates into reduced certainty about cause-and-effect relationships. In the Bounty paper towels experiment, all subjects would be exposed to both the fancy design and the plain design. It would be obvious that the research purpose pertains to the paper design, and any related measurements of a subject's attitude would be influenced accordingly. It would be difficult to conclude that differences in the responses of subjects were due to the experimental treatments. The differences could instead be due to a comparison effect where exposure to the one paper design results in the subjects comparing that design to the second. Also, subjects may give a more favorable response to the design type they feel is the most socially desirable. For example, subjects may wish to appear environmentally conscious and therefore express a preference for a functional (plain) paper towel. But in the actual marketplace, these same subjects may purchase the less functional (fancy) paper towel.

NOTE

1. Barry F. Anderson, *The Psychological Experiment: An Introduction to the Scientific Method* (Belmont, CA: Brooks-Cole Publishing Company, 1971), p. 28.

Chapter 3

Accuracy of Experiments

Marketing research projects are intended to provide accurate data about the topics and questions under study. These projects commonly involve statements expressed as hypotheses about cause-and-effect relationships of interest to marketing management. Experiments are generally acknowledged as the best research methodology for testing hypotheses about causal relationships and then making subsequent conclusions with reasonable accuracy. However, several aspects of an experiment are potentially detrimental to the accuracy of the collected data and the accuracy of the related conclusions. These aspects are experimental error, extraneous variables, demand characteristics, and validity.

EXPERIMENTAL ERROR

Many types of error can occur in an experiment. When error is referred to in an experiment without specific identification of type it is called experimental error. Therefore, experimental error is any type of error that occurs in an experiment. Its impact is a negative influence on the accuracy of the conclusions. Two common types of experimental error are random error and constant error.

While such errors are associated with experiments, many of them are rather generic since they also pertain to other research methodologies. For example, both random error and constant error are concerns for all research projects, as well as projects utilizing experiment methodology. Also, it might be noted that the term random error is often equated with the term sampling error and the term constant error is often equated with the term systematic error. While these respective terms are not completely synonymous, sampling error is a term used interchangeably with random erro and systematic error is a term used interchangeably with constant error.

Random Error

Random error occurs by chance. While it does not distort an experi ment's data in any particular direction, it does make the data less clear. I other words, its effect in experimental treatments is not consistent. If a experiment was repeated numerous times, a random error would effec different experimental conditions on different repetitions.

Random error is associated often with an experiment's sample of sub jects. Consider an experiment for the Kraft company for its Thousand Is land salad dressing. It could involve product composition as th independent variable, with two values: (1) "natural" with no preservative and a refrigerated life of three weeks after opening, and (2) "preserved with preservatives and a refrigerated life of six months after opening. Re search procedures in such an experiment would likely specify that the sam type of subjects be used, in the two experimental conditions, in terms o factors that possibly could impact preferences.

The list of factors that could impact preferences is long. For exampl three factors are: family size, product experience, and age. Subjects fro larger families may prefer "natural" because it's more healthy for the young children and since it is consumed quickly the short "shelf life" i the family's cupboard is not a concern. Subjects with product experienc also may prefer "natural" because they have experienced an unpleasa artificial taste with other preserved salad dressings. Subjects who are old may prefer "preserved" since these individuals may neither eat very muc salad dressing at a time nor eat it frequently. Therefore, their slow co sumption requires a product with longer refrigerator life to keep it fro spoiling before used. Otherwise, the product becomes stale and must b thrown away by these households before it is consumed.

Considering the possibilities and trying to assign certain subject types t certain experimental conditions is a difficult undertaking at best. To do s first requires determination of the significant characteristics. It then requir measurement of these characteristics in order to identify which people po sess them. To avoid these difficulties, to deal most effectively with th potential random error, the approach most preferred by researchers is ra dom assignment of subjects. With random assignment, experimental err related to different subject types is random error. As such, with rando assignment it is evenly distributed among experimental treatments. The r sult is that this random error may make the conclusions about cause-an effect relationships less clear, but at least it lessens the chance f conclusions in the wrong direction.

Constant Error

Constant error occurs systematically. Because its effect is consistent, an experiment's data are distorted in a particular direction. If an experiment was repeated numerous times, a constant error would effect the same experimental condition on different repetitions.

Constant error is more serious than random error. While random error leads to a less clear conclusion, constant error leads to a wrong conclusion. A reason is that constant error in an experiment is often due to extraneous variables known as confounding variables. A confounding variable is an extraneous variable that is not controlled by the researcher. The lack of control is because the researcher is not aware of the extraneous variable and therefore takes no action to eliminate it or to contend with it. As a result, the meaning of collected data is confounded or confused, which increases the likelihood that related conclusions are wrong.

Consider the above Thousand Island salad dressing experiment. If subjects from larger families were always assigned to the experimental treatment with the "natural" product, their responses in this situation would be consistently favorable. Since the marketplace is not comprised of only this type of subject, a conclusion by the Kraft company to produce, promote, and distribute only the "natural" version would be a wrong marketing decision.

EXTRANEOUS VARIABLES

Hypotheses specify the effect on a dependent variable caused by an independent variable. Variables outside of these hypothesized relationships known as extraneous variables can also cause an effect on a dependent variable. Formally defined, an extraneous variable (or confounding variable) is an uncontrolled variable that causes an effect on the dependent variable. These "extra" variables confuse (or confound) data about hypothesized relationships between independent and dependent variables. If extraneous variables cause or even partially influence the data collected in an experiment, subsequent conclusions and actions will likely be erroneous. Extraneous variables in an experiment to determine the effect of container color on product sales could be other aspects of the product (appearance, quality, reputation, package size, etc.), other marketing mix variables, and environmental variables (economy, competition, laws, technology, etc.).

An extraneous variable poses a threat to experiments. The threat is that researchers are interested in the effect caused by the controlled and manipulated variable under study, but other variables can confuse or confound the data. As a result, when conclusions are made about the effect on the dependent variable (such as product sales) caused by the independent variable (such as package color), these conclusions are likely wrong if othe factors (i.e., extraneous variables) caused or even partially influenced th effect measured.

Extraneous variables in marketing research always pose threats to th accuracy of conclusions based on experiments. When an experiment is con ducted in the marketplace with an aspect of the product variable as th independent variable (such as package color), extraneous variables that ca impact a dependent variable (such as that product's sales) are other aspect of the product (such as quality and package size), other marketing mi variables (promotion, price, and distribution), and uncontrollable environ mental variables (such as competition, economy, and societal attitudes).

The goal is to exert as much control as possible to minimize, if no eliminate, effects of these extraneous variables. A way to achieve this goa is to at least maintain the same influence of extraneous variables across al experimental conditions. When control is not possible, which is the situa tion with uncontrollable environmental variables such as competition an the economy, the best procedure is to be aware of the possible influence o these extraneous variables.

As Exhibit 3.1 indicates, conclusions about cause-and-effect relation ships, without careful attention to extraneous variables, are generally sus pect. Therefore, one approach to controlling extraneous variables is t minimize their effects by maintaining the same influence across all exper mental conditions. This control, depending on circumstances, can be fo tered through either random or matched assignment of subjects. Anoth approach, especially with environmental variables such as competition an economy conditions, is for the researcher and user of the research to simpl be aware of their possible effect, since recognition of a problem is ofte half the solution.

DEMAND CHARACTERISTICS

The problem of subjects responding in an experiment in a manner the think the researcher desires is just one dimension of potential demand cha acteristics that threaten the accuracy of an experiment. A demand chara teristic unintentionally provides subjects with information about the stud It consequently threatens the accuracy of data because subjects are like to respond differently when a demand characteristic exists than when does not. Instead of truthful responses, subjects tend to respond in a w they think researchers want or in a way they think is most socially desi able. The result ultimately leads to erroneous marketing decisions.

Demand characteristics are minimized by experiments that are proper designed and properly conducted. On the other hand, they exist when su jects know an experiment's hypothesis, independent variable(s), or oth information pertinent to the research purpose. Their knowledge or awar

Exhibit 3.1
Extraneous Variables: Reality Example

Pepsi and Michael Jackson

Marketing executives for Pepsi soft drinks were elated when singer Michael Jackson agreed to be their spokesperson. However, because of his popularity, his endorsement was expensive. While such contracts are normally confidential, Pepsi reportedly paid Michael Jackson $15 million at the time of his "Thriller" album, as much as $45 million at the time of his later "Bad" album, and up to $90 million was reported at the time of his "Dangerous" album.

With such high expenditures, the marketing executives conducted an experiment to justify these amounts to the company and their stock holders. The independent variable was the spokesperson program for Pepsi, the independent variable values were the program (1) with Michael Jackson and (2) without Michael Jackson. The dependent variable was sales, and the hypothesis was that Michael Jackson as a Pepsi spokesperson causes substantially higher Pepsi sales.

Results of the experiment indicated that Pepsi sales were higher with Michael Jackson as a spokesperson. In fact, Pepsi sales in the entire country of Japan doubled during the singer's "Bad" concert tour of that country. However, such conclusions about causal relationships are seriously flawed since numerous extraneous variables, that could have caused the increase in sales or at least influenced them, were not controlled or even considered.

For example, during Jackson's most intense endorsements, price was lowered, which placed Pepsi on special sales, distribution was adjusted to double stock the channels of distribution, and promotion included increased point-of-purchase displays, contests, and publicity. At the same time, changes in the marketing environment included increased press coverage, increased media exposure of Pepsi, and an actual temporary cutback of marketing expenditures by the competition. All or any of these factors (i.e., extraneous variables) no doubt caused or contributed to the increase in sales, making it impossible to conclude with reasonable confidence that the increase in Pepsi sales was caused solely by the endorsement by Michael Jackson.

ness can come from speaking with other subjects, personal cues from a researcher such as nonverbal body language and variations in tone of voice while giving instructions, and indiscreet experimental procedures. Demand characteristics also can occur simply by the subjects being aware that an experiment is being conducted.

Some demand characteristics are very subtle. As Exhibit 3.2 reports, one type of demand characteristic known to occur is when the only information the subjects have is an awareness that they are involved in a research study. Originally discovered in regard to a management experiment, it is now commonly known as the Hawthorne Effect.[1] The Hawthorne Effect is a

Exhibit 3.2
Hawthorne Effect: One Type of Demand Characteristic

```
One demand characteristic was identified early and
labelled as the Hawthorne Effect. According to the
Hawthorne Effect, subjects behave differently than normal
when simply aware they are part of an experiment.

Identification of the Hawthorne Effect emerged from the
original "Hawthorne studies" to investigate optimal
conditions for productivity of factory workers. When
lighting was investigated, experimental data revealed that
employee productivity increased regardless of which of
three experimental treatments that a subject received:
(1) increased lighting, (2) decreased lighting, and (3)
status quo lighting. Researchers concluded that increased
productivity in all these three conditions was due to
awareness on the subject's part that they were being
studied.

______________________

Additional information about the Hawthorne Effect can be
obtained in contemporary university textbooks in
organizational behavior and human resource management, or
by consulting the original report written by F. J.
Roethlisberger and W. J. Dickson, Management and the
Worker, Cambridge, MA: Harvard University Press, 1939.
```

type of demand characteristic in which subjects behave differently than normal simply because they are aware they are part of an experiment.

Consequences

Demand characteristics ultimately have negative financial consequences. These consequences occur because corresponding marketing decisions are based on inaccurate information. Consider a potential experiment with Cheerios breakfast cereal, marketed by General Mills. The independent variable is package size, with two values: current 1-pound size and proposed 3-pound size. The hypothesis is that consumers will prefer a 3-pound size more than the current 1-pound size. To illustrate the negative effect of demand characteristics, assume:

1. that General Mills' marketing executives (because of greater product sales, and less production, inventory, and distribution costs) hope consumers prefer the 3-pound size,
2. that subsequently, albeit unintentionally, they will communicate this hope to the research participants, and

3. that although subjects prefer the 1-pound size (e.g., to avoid spoilage and to fit into home cupboards), they indicate preference for the 3-pound size during the experiment because, in their minds, this is what management wants, and it is what a "smart shopper" would prefer, since a larger size would cost less per ounce.

Given this hypothesis, that consumers will prefer a 3-pound size over the current size, marketing executives at General Mills will hope consumers prefer the 3-pound size, since it means more sales. In fact, since marketing the 3-pound size could offer a financial benefit to the company, this would be the major motivation by company management to approve marketing research to test the hypothesis. The financial benefit begins with less production costs and less inventory costs, especially if the 1-pound size is discontinued. Distribution costs would also be lower to market the product in only the larger package size. But given these strong financial benefits, it is important that management's hopes regarding the consumers' preferences do not translate into demand characteristics. In order to have valid data upon which to make this package size decision, it is necessary that the research hypothesis and interests of management not be communicated to the subjects.

Actual preference of the subjects may be for the smaller 1-pound size (e.g., to avoid spoilage and to fit into their home cupboards). To comply with these demand characteristics, the subjects may indicate a preference for the 3-pound size during their involvement in the experiment, because this is what management seems to want. At the same time, it is what a "smart shopper" would select, since a larger size could be expected to cost less per ounce than a smaller size. However, despite results of the research data, since smaller package size is in reality preferred, the larger size would not be purchased in the marketplace.

Data from this experiment, if influenced by demand characteristics, are likely to lead to an erroneous conclusion and subsequent erroneous decisions by management: that the large package is favored, followed by a decision to discontinue the smaller package size and replace it with the larger package size. The result might be a huge financial cost to the company, as consumers in the actual marketplace choose a competitor's breakfast cereal which is marketed in the desired smaller package size. A similar situation occurred a few years ago with the Coca-Cola company in the soft drink market as Exhibit 3.3 indicates.

Control

Researchers can control demand characteristics. The control starts at the planning stage of a research project. Aspects of an experiment that could produce demand characteristics can be and should be identified with

Exhibit 3.3
Demand Characteristics at Coca-Cola: A Reality Example

After conducting many experiments, marketing researchers at the Coca-Cola corporate headquarters in Atlanta, Georgia, concluded that a definite preference exists in the soft drink market for "new Coke" over "old Coke." This conclusion was based on data from subjects in parts of Texas where subjects in a 3-to-1 ratio preferred new Coke over old Coke, and in Detroit the ratio was as high as 8-to-1.

Based on the marketing research, corporate executives made the subsequent decision and implemented the necessary actions to discontinue old Coke and market only new Coke. But purchases made in the marketplace did not parallel preferences expressed in the experiments. Once the related marketing actions were taken by Coca-Cola management, consumers expressed uproar to the extent of boycotting new Coke. Shortly thereafter, and with huge financial consequences, Coca-Cola began to again market old Coke, now called Classic Coke.

The reason that reality was not consistent with research data, and erroneous corporate decisions were made, could have been an influence by demand characteristics in which research participants felt compelled to tell the researchers what they thought the researchers wanted to hear: that new Coke was superior in taste to old Coke.

thorough advance planning. Identification is again "half the battle." Once potential demand characteristics are identified, procedures can be incorporated at the planning stage (i.e., research design stage) to eliminate them. Accomplishing this task involves scrutinizing every aspect of the experiment. Part of this scrutiny is to rehearse the research personnel who will interact with the subjects. The goal here is to ensure that no nonverbal cues provide unintended information.

To control demand characteristics, disguise is often necessary. However, it raises such questions as: Should control be exercised to the extent that subjects are misled or even lied to? If so, is it illegal and will they be injured? This form of disguise can involve questions of legality, most evident in situations where subjects are in some way injured. The answer here is to avoid this extent of disguise; even if it is necessary to avoid the entire experiment. If the disguise is legal, another series of questions pertains to the ethics of the situation: Even if disguise in this experiment is legal, the researcher must ask if it is ethical. Answers to the legal questions may be clear-cut, but answers to the ethical questions depend on a researcher's personal ethics, respective professional organization code of ethics[2] (e.g.,

Exhibit 3.4
Selected Values Identified by Professional Organizations Associated with the Marketing Research Profession

Information presented here demonstrates that the intention of the Code of Ethics for organizations associated with the marketing research profession is for researchers to treat marketing research participants fairly and with respect.

Associations	Nondeception	Serving Others	Not Harming Others	Justice
	Values Specifically Identified by Respective Codes			
Advertising Research Foundation	X			
American Marketing Association	X	X		X
Council of American Survey Research Organizations	X	X		
Marketing Research Association	X	X	X	
Qualitative Research Council of America		X		

This table is based on a working paper by Stephen B. Castleberry and Warren French (1991), "The Ethical Framework of Advertising/Marketing Research Practitioners: A Moral Development Perspective;" as reported by Patrick E. Murphy and Gene R. Laczniak (1992), "Emerging Ethical Issues Facing Marketing Researchers," Marketing Research: A Magazine of Management and Applications, vol. 4, no. 2, June, pp. 6-11.

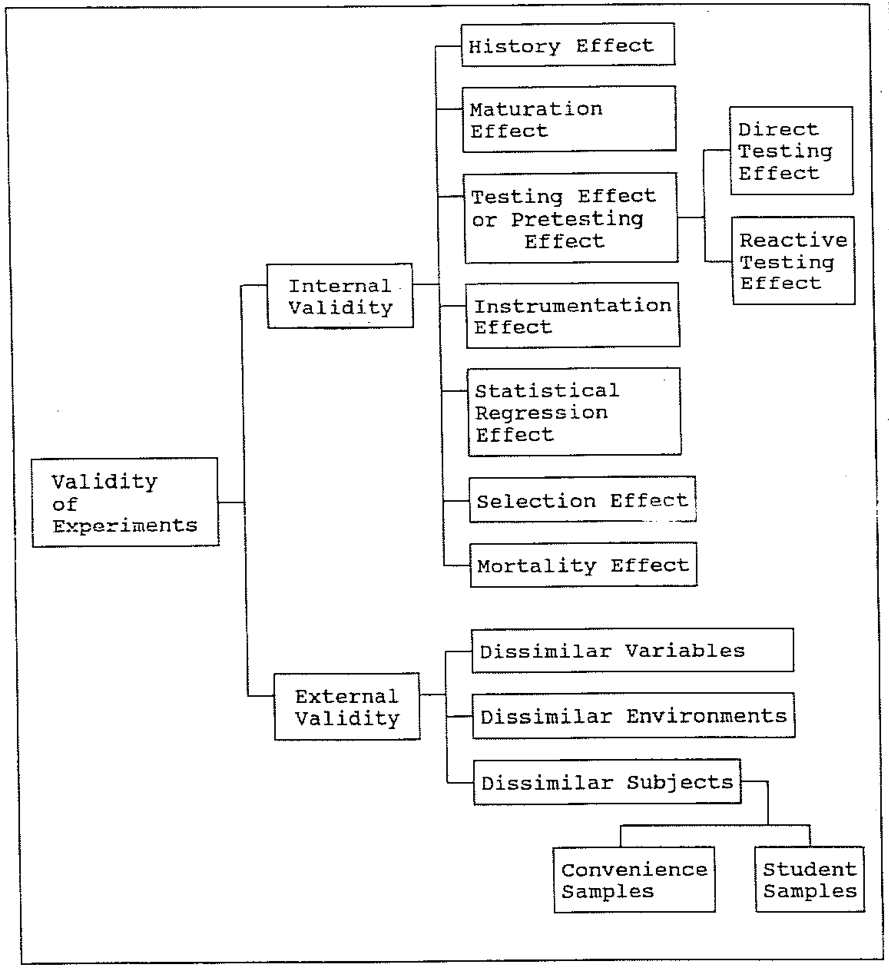
see Exhibit 3.4), and/or guidelines established by a particular company management.

A pertinent research procedure to deal with disguise is with a debriefing—a postexperiment activity in which the research purpose and the subject's contribution are explained, and any questions which a subject may ask are answered. Researchers usually meet this responsibility in the form of a short meeting, conversation, or printed handout given to subjects after they complete their participation in an experiment.

VALIDITY

Validity is the extent to which a measurement is free of error in terms of providing information intended.[3] Its complement is reliability—the ex-

Exhibit 3.5
Dimensions of the Validity of Experiments



tent to which a measurement is free of error in terms of providing consistent information. Because an experiment can have reliability without validity, but not validity without reliability, it is appropriate here to consider validity in slightly greater detail.

As Exhibit 3.5 shows, two major dimensions of validity are internal and external. Internal validity is the extent to which changes in a dependent variable are actually caused by changes in an independent variable. It is increased by eliminating the influence of extraneous variables. External validity is the extent to which the measures (i.e., collected data) in a particular research project can be generalized.[4] It is increased by confirming answers

to such questions as: Does a particular experiment reflect the reality of the surroundings outside the experiment? Do the results apply to the larger population of interest? Do the results of the experiment help us to understand the larger world?

Internal and external validity are not mutually exclusive. It is usually necessary to forego some internal validity to increase the reality (i.e., external validity) of an experiment. For example, external validity is increased by experiments that reflect reality. But reality implies a naturalistic setting that is complete with more variables than can be controlled. In contrast, internal validity is enhanced by controlling the entire experiment, which requires a small number of variables in a setting much different than reality. Thus, to achieve reasonable levels of both internal and external validity it is commonly necessary to make trade-offs.

NOTES

1. F. J. Roethlisberger and W. J. Dickson, *Management and the Worker* (Cambridge, MA: Harvard University Press, 1939).
2. The interested reader should see an article by Patrick E. Murphy and Gene R. Laczniak, "Emerging Ethical Issues Facing Marketing Researchers," *Marketing Research: A Magazine of Management and Applications* 4, no. 2 (June 1992), pp. 6–11.
3. See earlier, well-accepted material presented by Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1966), pp. 5–6.
4. In recent years, some researchers have begun to use the term "ecological validity" to describe this focus on the larger world outside a particular experiment. Although not yet a frequently used term, ecological validity is synonymous with external validity and, in reality, the two terms are used interchangeably.

Chapter 4

Internal and External Validity

Extraneous variables, experimental error, demand characteristics, and validity are all important dimensions of the accuracy of experiments. However, validity is of particular concern, since it reflects the extent to which data are influenced in a negative manner in a particular experiment because of how the experiment is conducted. In other words, validity is often impacted directly by an experiment's extraneous variables, experimental error, and demand characteristics.[1]

As defined in Chapter 3, validity is the extent to which a measurement is free of error in terms of providing information intended. Validity is reflected in the data collected for an experiment. If data lack validity (i.e., are not free of error and/or do not represent what was intended to be measured), conclusions based on these data are not justified (i.e., not valid). When these conclusions are not justified, related management decisions and subsequent actions are likely to be erroneous and correspondingly costly to the organization.

Validity is an issue with which personnel who conduct marketing research and those who make decisions based on the results must be concerned. Concern about validity applies to the times before, during, and after an experiment is conducted. Consider the above General Mills experiment to test the hypothesis that a larger package size of Cheerios breakfast cereal will cause more product sales. What is intended to be measured is the relationship between package size (independent variable) and product sales (dependent variable). An experiment with good validity will yield error-free data that actually measure this relationship. These data then encourage research conclusions that are justified, followed by prudent marketing decisions and actions by the company management.

Validity is multidimensional. There are different types reflecting different concerns. As the formal definition indicates, one aspect of validity focuses on mistakes that occur in the measures (i.e., extent to which they are free of error) and a second aspect that focuses on what the measures actually measure (i.e., extent to which what is measured is intended to be measured). These two focuses reflect the primary distinction between types of validity identified as internal validity and external validity. Internal validity is the extent to which changes in a dependent variable are actually caused by changes in an independent variable. It is increased by eliminating the influence of extraneous variables. External validity is the extent to which the measures (i.e., collected data) in a particular research project can be generalized.

Internal and external validity are not mutually exclusive in experiments. The extent of generalization that is possible depends on how realistically the experiment is conducted. But to increase the reality of an experiment it is often necessary to forego some internal validity. For example, external validity is increased by experiments that reflect reality, and this requires a naturalistic setting that is complete with a multitude of variables. In contrast, internal validity is enhanced by tightly controlling the entire experiment, and this requires a relatively small number of variables in a setting that is typically much different than reality. Thus, to achieve reasonable levels of both internal and external validity, it is commonly necessary to make some trade-offs.

INTERNAL VALIDITY

Internal validity is decreased to the extent by which variables other than the independent variable (i.e., extraneous variables) can effect or account for changes measured in a dependent variable. At least seven types of extraneous variables pose threats to internal validity through their respective effects: history, maturation, testing, instrumentation, statistical regression, selection, and mortality.

History Effect

History is the elapse of time during which changes occur. A history effect is the effect on a dependent variable by an extraneous variable associated with the passing of time. Its relevance is that experimental procedure typically measures a dependent variable at one period of time, exposes subjects to an experimental treatment, and then measures the dependent variable again at a later period of time. The related history effect occurs between the two measurements. Since the exposure to a treatment occurs concurrently with the passage of time, it can be difficult to determine if the treat-

Exhibit 4.1
Deciding to Market Coca-Cola by the 3-Liter: Potential History Effect

An experiment utilized in a test market with a company such as Coca-Cola can be readily susceptible to history effect. Consider an experiment that involves questions about the company's decision to market Coke in a new 3-liter size; in addition to the plastic take-home bottles already sold in 2-liter and 1-liter sizes.

The experiment involves a test market in eight locations (frequently selected for test markets): San Diego, Los Angeles, Cincinnati, Kansas City, Chattanooga, St. Louis, Baltimore, and Philadelphia. Its independent variable is container size for Coke and the dependent variable is total volume of Coke sales. Total sales volume of Coke is first measured in each of these cities, followed by introduction of the 3-liter container size for three months in four of the cities. At the end of the experimental treatment period, a second measure of total sales volume of Coke is measured in each of these cities.

Assuming significant changes in sales volume has occurred in the locations with the 3-liter containers, they could be due to a history effect. For example, during the three months of time that elapsed during the experiment, the economy could have changed (such as higher inflation which may influence customers to purchase larger quantities), the competition could have decreased its marketing efforts, and the weather could have been unseasonably warm. Therefore, it is not clear if a change indicated by the second measure can be reasonably attributed to the independent variable manipulation of container size, or if the change was caused by some history effect that occurred during the time the test market was conducted.

ment or time accounts for changes indicated between the two measurements.

The history effect is especially troublesome with experiments used to conduct a test market. They, in effect, pose a classic "Catch-22" situation. If a test market is run for a longer period of time, more information is collected. But a longer time means greater opportunity for other events, external to the experiment, to impact the outcome (e.g., see Exhibit 4.1). Despite this fact, it is not unusual for major companies to conduct test market experiments that span a year or more. Within the three-plus years of research that McDonald's conducted for its pizza, a test market was conducted for over a year before making a decision about marketing the product throughout its organization. Caution must, of course, be exercised in interpreting these results. During such a lengthy experimental treatment

time for the test market, history effects can be in the form of any of the variables uncontrollable to marketing (ranging from changes in the economy to changes in the marketing efforts of competitors).

Maturation Effect

People change—psychologically and physiologically—as they age. College students certainly change as they mature from 18-year-old high school graduates to when they graduate from college. Organizations and other potential subjects in an experiment also change as they age. These changes can pose problems in marketing research known as maturation effects. A maturation effect is the effect of an extraneous variable associated with the normal aging process of a test unit. This effect is closely related to the history effect, and can be considered a type of history effect that focuses on the test units/subjects.

Even though changes occur gradually and subtly during the normal aging process, the longer an experiment is conducted, the more likely the results will be influenced by the maturation effect. Not all experiments are affected in the same way, since these changes and influences differ among groups of people. For example, changes are more rapid and visible when the subjects are very young children, adolescents, and very old.

To realize this effect, consider a two-year experiment for a new Clearasil acne medication. Since teenagers represent the target market, subjects would be teenagers in their "prime acne years." However, the maturation effect would severely influence an experiment of this length with these subjects. The reason is that teenagers change considerably during two years as they mature and outgrow their acne. While this example represents an extreme situation to illustrate the maturation effect in experiments, all experiments are susceptible to the maturation effect, though often at a level that is much more difficult to discern.

Testing Effect

Experiences change people. A person is never exactly the same after being exposed to an experience. Even though the effect may be minute, people are likely to respond differently to a stimulus they are exposed to for the first time compared to the second time. When assessing measures in an experiment, this fact poses problems referred to as testing effects.

A testing effect (or pretesting effect) is the effect on a dependent variable by an extraneous variable associated with exposure to the measure of the dependent variable. Merely being exposed to a measure of the dependent variable changes response to a second measure, regardless of whether the experimental treatment is administered or not. Furthermore, being exposed

Exhibit 4.2
An Example of Testing Effect: An Experiment to Test Advertising Appeals for Clearasil Facial Clean

> Consider an experiment regarding advertising (as the independent variable) for Clearasil Facial Clean, a new acne medication. Company management wants to determine which of three advertising appeals is best to generate interest to purchase the product (dependent variable) among a target market comprised of 16-year-olds. The three appeals under consideration:
>
> (1) Immediate appearance appeal—message that using Clearasil Facial Clean makes facial skin look more healthy immediately.
>
> (2) Immediate health appeal—message that using Clearasil Facial Clean will not make facial skin look more healthy immediately, but will immediately make the skin more healthy medically.
>
> (3) Long-term health appeal—message that using Clearasil Facial Clean will not in any way benefit the appearance of facial skin in the short term, but will in the long term make the skin more healthy and the appearance better, after a person passes the prime acne years and goes into his/her twenties.
>
> To determine the effectiveness of these appeals with an experiment, it is necessary to measure <u>initial</u> interest to purchase the product (i.e., obtain a baseline measure against which a second measure is compared). After the initial measure, different subjects are then exposed to the advertisements, followed by a second measure of interest to purchase. However, subjects in this situation may have never even considered the long-term health consequences of an acne product. But due to testing effects, the first measure made them aware of the issue. In turn, this awareness could cause completion of the second measure to be different than if awareness was not raised by the first measure. It could also increase sensitivity to the advertising appeals, through the dynamics of selective perception.

to such a measure produces a different response to the experimental treatment compared to not being first exposed to the measure.

Attempting to eliminate testing effect makes planning an experiment very difficult, as suggested in Exhibit 4.2. The reason is that, in order to most accurately determine the effect caused by the independent variable, it is necessary to first measure the dependent variable before subjects are exposed to an experimental treatment. However, this first measure provides subjects with a different experience than if never exposed to it. At the least, it provides more acquaintance with the specific measures, possibly suspicion about the experiment's intent, and likely increased sensitivity to the eventual experimental treatment. The result is that any change indicated by the second measure of the dependent variable (after subjects are exposed to the experimental treatment) could be due partially to a testing effect rather than solely to the effect of the experimental treatment. Because the effect occurs before subjects are exposed to the experimental treatment, it is also referred to as a pretesting effect. Therefore, pretesting effect is synonymous with testing effect and, in reality, the two terms are used interchangeably.

Two types of testing effects occur—direct and reactive—and both are illustrated in the acne product example (Exhibit 4.2). Direct testing effect (or main testing effect) is when the first measure in an experiment impacts the second measure due to experience with the measures. Thus, even with no exposure to an experimental treatment, subjects involved in the first measure (e.g., completing a pretest questionnaire) complete the second questionnaire differently than if they had never done the first questionnaire. Reactive testing effect (or interactive testing effect) is when the first measure in an experiment impacts the second measure by sensitizing the subject to the experimental treatments. By completing the first questionnaire, subjects are alerted or sensitized to the following experimental treatment. In the Clearasil experiment, subjects are possibly alerted for the first time ever to long-term consequences of acne products. As a result, they react differently to the advertising appeals than if they had not been alerted to the issue.

In reality, more than one term is used when referring to these testing effects. For practical purposes, the term "main testing effect" is synonymous with direct testing effect and, in reality, the two terms are used interchangeably. Also, the term "interactive testing effect" is synonymous with reactive testing effect and, in reality, the two terms are used interchangeably.

Instrumentation Effect

Sometimes the questionnaire or other measurement device used to assess a dependent variable is changed. Since the means used to assess the dependent variable was historically considered a tool or instrument, effects due to related changes are known as instrumentation effects. Instrumentation effect is the effect on the dependent variable, caused by a variable external to an experiment, which is associated with a change in the measurement instrument. It is consequently difficult to conclude that an apparent difference in measurement results is caused by different experimental treatments or by different measurement instruments.

An instrumentation effect is likely to occur any time the measurement instrument is not identical for all measures, including when an individual subject is measured more than once and when the instrument is administered to more than one subject. For an individual subject, a change in the measurement instrument occurs between the first and second measures of the same subject. For more than one subject, changes in the measurement instrument occur between measures of one subject and another subject.

Instrumentation effects occur readily with marketing research. Four situations common to marketing research, that can lead to instrumentation effects, are when:

1. Questions and words in a questionnaire are added, omitted, or reordered during a research project.
2. The same researcher/interviewer interacts differently with either different subjects or the same subjects at different times. It occurs because researchers are human, their energy, boredom, personal problems, appearance, dress, and liking and comfort with different subjects all differ during a work shift. The consequent can be significant differences in tone of voice, speed of speech, body posture, appearance, friendliness, politeness, patience, demeanor, and attitude.
3. Although the questionnaire is the same, multiple interviewers/researchers may administer it. The potential for problems is that these multiple researchers/interviewers interact differently with subjects. These differences are natural because people are not identical. However, these differences multiply differences identified above in point #2 regarding the same researcher.
4. Measurement instruments in marketing research are not limited to questionnaires. Mechanical-electronic devices are used often, and calibration of these devices changes, due either to altering by a researcher or by natural wear. Either change means that data which indicate an apparent impact due to experimental treatments may instead be due to a change in the measurement device. Exhibit 4.3 illustrates such a situation.

Statistical Regression Effect

Extremes tend to average out over time. It occurs in the natural sciences and social sciences, as well as in marketing and marketing research. For example, the field of biology shows that over time, children of tall parents and children of short parents tend to produce children of more average height. The fields of sociology, political science, and psychology show that strongly held attitudes tend to also move toward the average (i.e., moderate) over time. In experimental research these changes from the extremes are known as statistical regression effects. A statistical regression effect is an effect of an extraneous variable, on the dependent variable, by an extraneous variable associated with a subject characteristic that moderates from its extremes.

These effects can confound marketing research data collected by experiments. To begin, a statistical regression effect with negative consequences is likely to occur when subjects are assigned to experimental treatments

Exhibit 4.3

An Example of Instrumentation Effect Involving Research in the Television and Advertising Industries: A Firsthand Experience Example

To illustrate instrumentation effects, consider a rather standard testing procedure involving mechanical-electronic devices for marketing research in the television and advertising industries. The project might be an experiment to test two new potential television series for the CBS Television Network: "Home Sweet Home," a situational comedy and "Double Dare," a detective-mystery. The dependent variable is feelings (liking and disliking) that subjects experience while watching an episode of these programs.

To measure these feelings, an electronic-mechanical mechanism is used to assess the dependent variable. The mechanism has two push-buttons: a green button for "feelings of liking" during the show and a red button for "feelings of disliking." Every second of the television program or advertisement a computer records which button is pushed by which subjects.

Numerous instrumentation effects can occur:

(1) An obvious change is if researchers alter the computer recording from one second to two seconds.

(2) A subtle change occurs as the push-button mechanism wears out. Therefore, the spring-activated pressure necessary to push a button changes slightly during a one-hour test show, so that it is more difficult to register a particular feeling at the beginning of a program (when the spring is most firm) than later in the show. Also, substantial change occurs during a year's use, such that the same program tested early in the year can indicate different feelings than later in the year, caused <u>not</u> by different audience feelings but by changes in effort necessary to register a feeling.

(3) As new spring-activated devices arrive from the factory, the pressure required to operate them may vary. The consequent is that two subjects seated beside each other, using new devices, may express the same feelings toward a scene in a test program, but their responses could be measured or recorded differently if the device button is not pressed sufficiently hard.

based on pretreatment measures. With this assignment procedure the strength of a relevant subject characteristic is assessed before the experiment begins. Subjects identified as high/strong and low/weak on this initial assessment are assigned to experimental groups accordingly. The problem is that subjects identified with the extreme levels of the characteristic used to assign the subjects tend to regress or move back to more moderate levels over time. Therefore, a change occurs in groups of subjects between the first and second measures of an experiment, during the same time the experimental treatment is administered. The data are consequently confounded, meaning that it is unclear if the results are due to statistical regression effects and/or experimental treatment effects.

Since marketing research experiments are susceptible to the statistical regression effect, caution must be exercised when using an extreme level of some characteristic to assign subjects to groups. Consider the earlier experiment for Clearasil Facial Clean. Assume subjects were assigned to experimental treatments based on completed questionnaires that identified acne concern as either high or low. Individuals highly concerned were likely those with acne at the time of the initial questionnaire. But as their acne lessened, during the time they naturally passed through their prime acne years, their concern no doubt lessened accordingly. As a result, conclusions about lessening acne concern, caused by the experimental conditions, are unclear, since such change would be also an expected statistical regression effect.

An experiment such as this one is very susceptible to the statistical regression effect. It is especially likely as the time to conduct the experiment is increased. The problem is that the concern of these two groups of subjects is likely to change naturally as they pass through their prime acne years. As a result, level of concern would no longer be a meaningful characteristic that accurately identifies the same subjects by the time the posttreatment measure is assessed.

Selection Effect

It is generally undesirable to have a disproportionate number of any one type of person assigned to the same experimental treatment. The reason involves selection effect—the effect, on a dependent variable, by an extraneous variable associated with different types of subjects not being evenly distributed between experimental treatments. The problem is the tendency for the same type of subjects to respond similarly, irrespective of exposure to experimental treatment.

Selection effects that result in uneven distribution can occur through self-selection by subjects, as Exhibit 4.4 indicates, or systematic assignment by researchers. The solution to avoiding both these occurrences is random assignment. If a large number of subjects possess a certain characteristic, random assignment will distribute these individuals evenly between the experimental treatments.

Mortality Effect

People regularly discontinue their involvement in clubs, activities, relationships, and so on. The reasons are many, and include people changing their minds, becoming too busy, moving, ultimately dying, and so forth. Discontinuing involvement is natural, but in experiments it has a detrimental impact in the form of mortality effects. A mortality effect (or sample attrition) is an effect on the dependent variable, caused by an extraneous

Exhibit 4.4
An Example of Selection Effect: An Experiment to Test Advertising Appeals for Clearasil Facial Clean

```
Consider again the experiment for Clearasil Facial
Clean. The independent variable is advertising appeal,
and the independent variable values are (1) immediate
appearance, (2) immediate health, and (3) long-term
health. The dependent variable is attitudes toward the
product.

A way to obtain subjects for this experiment is to
advertise in high school newspapers for volunteers.
However, a selection effect (i.e., self-selection) with
this approach is that volunteers likely will be
sufferers of acne and, in addition, probably only those
who are most bothered by it.

The best approach generally agreed on by researchers is
to randomly assign individual subjects to the
experimental treatments. Therefore, a researcher might
arrange first with a school administration to randomly
select homerooms, and then second, randomly assign
individual subjects in these homerooms to one of the
three experimental treatments.
```

variable, that is likely to occur when subjects withdraw from an experiment before it is completed.

If a substantial number of subjects withdraw between the first measure and the second measure of an experiment, changes in the dependent variable will be difficult to attribute solely to manipulations of the independent variable. Because the mortality effect involves subjects withdrawing from the experiment, it is also referred to as a sample attrition effect. Therefore, sample attrition effect is synonymous with mortality effect and, in reality, the two terms are used interchangeably.

Mortality effects make it difficult to determine changes both within and between experimental treatments. An excessive number of subjects withdrawing within a particular experimental condition makes it difficult to compare changes between the first and second measure. These withdrawals also make it difficult to compare changes between experimental conditions when one experimental condition experiences more subjects withdrawing than another.

Factors such as an experiment's length of time, type of sample, type of experimental treatment, and other factors as indicated in Exhibit 4.5 all contribute to mortality effects. The longer an experiment is conducted, the more likely subjects will withdraw. Withdrawal may be due to aging and dying, moving to another city, or personal lack of time to be involved in an experiment. The problem of aging and dying is an especially difficult

Exhibit 4.5
An Example of Mortality Effect: An Experiment to Test Product Versions for Clearasil Facial Clean

Mortality effects are a problem for experiments that require a subject's participation during an extended period of time and for experiments that have a treatment/condition that subjects find unpleasant.

Consider an experiment for a revolutionary new Clearasil Facial Clean product. The experiment is conducted for six months, from April through September. Subjects are teenagers at a local high school. The independent variable is a product with two versions (i.e., values):

(1) a product that leaves a transparent film on the face after each use and takes 90 days to improve acne, and

(2) a product that leaves an obvious film on the face after each use, but takes only 30 days to improve acne.

After the respective 90 days and 30 days, a second product must be used continually to maintain the improvements of each. The dependent variable is attitude toward the product.

Many mortality effects are likely in this experiment. Since it runs six months from April through September, some students will withdraw when parents accept job transfers to another part of town or another city and others will withdraw because of the long duration. Also, embarrassment will cause some individuals to withdraw, especially subjects assigned to the second experimental condition with the obvious film that remains on the face.

problem when dealing with elderly subjects. On the other hand, lack of personal time is an especially difficult problem when dealing with young mothers raising a family and working, as well as busy executives.

EXTERNAL VALIDITY

Do the results of a particular experiment help us to understand the larger world? Do the results of a particular experiment apply to the larger population of interest (from which the sample subjects were drawn)? Do the results of a particular experiment pertain to similar situations but in different settings? These are questions of external validity. As defined earlier, external validity is the extent to which the measures (i.e., collected data) in a particular research project can be generalized. It is increased and de-

creased by how well an experiment reflects the reality of the world outside the particular experiment.

External validity is desirable, but not at the exclusion of other important aspects of an experiment. For example, as experiments are conducted in a manner that increasingly reflects reality, research control is given up. This lessening of control poses negative implications for internal validity. Therefore, experiments should be conducted with a balance between internal validity and external validity. The proper relationship must be achieved between the internal components and procedures of an experiment and the larger world outside of the experiment. In recent years, some researchers have begun occasionally to use the term "ecological validity" to describe this focus on the larger world outside a particular experiment. Although not yet a frequently used term, ecological validity is synonymous with external validity and, in reality, the two terms can be used interchangeably.

In summary, the more an experiment differs from the world outside that experiment, the less its external validity for the subsequent data, results, and conclusions. Prominent factors that lessen external validity are variables, environments, and subjects, that are dissimilar between an experiment and reality.

Dissimilar Variables

The quantity and type of variables controlled in an experiment are always dissimilar to reality. Even sophisticated experiments are limited to relatively few independent variables, each with a few values, and often of a type that only approximates reality.

Consider the experiment for Clearasil Facial Clean. Its objective is to test which of three advertisements generates the most purchase interest: immediate appearance condition, immediate health condition, or long-term health. To accommodate needs related to internal validity, subjects are brought into a room where all factors (e.g., lighting, seating, noise, etc.) are controlled (i.e., held constant) for each subject. But external validity is limited because the quantity and type of variables that influence actual purchase decisions are very dissimilar.

While this experiment might control two variables with limited values (two product versions and three advertisement versions), the number of potential variables combined with their variations is almost limitless. They begin first with an almost limitless quantity of uncontrollable marketing variables (e.g., competition, economy, technology, societal attitudes toward acne, and so forth) each of which also has an almost limitless quantity of variations.

Second, a large number of controllable marketing variables also are not acknowledged in the experiment. For example, marketing management has many options not investigated in this experiment: commercial length (15 seconds, 30 seconds, 60 seconds, and even infomercials ranging anywhere from 5 minutes to 60 minutes), commercial type (with different appeals such as safety, convenience, value, social acceptance, etc.), delivery volume of commercial, stores where the product is sold (e.g., pharmacies, discount stores, department stores, grocery stores, health food stores, etc.), price, package size and design, publicity surrounding the product, pertinent consumer behavior (influence of friends and family), and so on. In addition, since television advertisements are expensive to produce, mock-ups are presented to subjects in the form of a storyboard or rough cut (unfinished) videotape.[2] However, such test material (advertisement mock-ups) is nowhere near the same as finished advertisements of broadcast quality.

Dissimilar Environments

The marketplace is difficult to duplicate, especially with the competing interest to maintain reasonable internal validity. The result is an environment or setting in an experiment that is often very dissimilar to reality, ranging from very similar at best to very dissimilar at worst. Toward the similar end of the continuum are experiments conducted as test markets in the actual marketplace. Toward the dissimilar end are highly controlled experiments conducted in laboratories. While these latter marketing research experiments often provide valuable information, they are almost always low in external validity.

Consider an experiment by General Mills, to decide the best package design (photos, layout, and colors) to market a new, improved Wheaties breakfast cereal. The dependent variable is attention paid to the box, measured by the length of time subjects focus their eyes on parts of the box. To conduct this measure, subjects are seated in a special chair while their heads are immobilized and two eye cameras attached to a futuristic-looking apparatus are positioned near their faces.

This type of experiment has good internal validity in that manipulation of the independent variable (such as placement of different photos on different package layouts) affecting the dependent variable (i.e., focus on different package aspects) is limited to manipulations of the independent variable. However, external validity is likely very low, due to the very artificial environment in which the experiment is conducted. For example, the package is viewed now in an environment completely different from a typical shopping trip, where the consumer is pushing a grocery cart, occasionally bumped, often distracted, interrupted by other shoppers, possibly in a hurry, surrounded by various background noises, and confronted with a kaleidoscope of competing products and packages.

Dissimilar Subjects

Consumers reflect the heart of marketing, and subjects reflect the heart of marketing research experiments. The data obtained from subjects form the basis upon which conclusions about subjects in the marketplace are drawn and subsequent marketing management decisions are made. Therefore, the more similar subjects are to consumers, the more research results may be generalized from an experiment to the marketing manager's actual decision situation. When subjects are substantially different from consumers who comprise the population of interest to marketing decision makers, generalization is questionable. As a result, the appropriate type of subjects to represent consumers and customers of interest to the marketing decision makers, and, correspondingly, the proper procedures and considerations regarding samples and sampling, are vital to achieving external validity in marketing research experiments. Otherwise, external validity is lowered.

The more similar the subjects of an experiment are to consumers, the more readily the results may be generalized. However, while it is desirable to have subjects identical or highly similar to the population of interest, it is not always imperative. Depending on the situation, external validity is not necessarily breached when subjects are different from marketplace consumers. Related considerations involve convenience samples and student samples.

Convenience Samples. One situation in which subjects are not identical to members of an actual target market involves convenience samples. A convenience sample is a sample comprised of subjects who are readily available. Examples are a boy scout troop, a class of university students, a ladies' church group, or some other easily obtained group of people that are recruited to serve as subjects. These types of convenience samples are generally acceptable if the subjects are serving in an experiment for a product of interest to them. While such subjects are likely to be more homogeneous in some characteristics than the population as a whole, relevant demographics and psychographics can be quite similar to the target market to which generalizations are intended to be made.

Caution must be exercised, but convenience samples are not automatically a negative influence on external validity. Their use is quite acceptable when the dependent variables and/or the intended generalizations are compatible with the population of interest. In other words, they are acceptable when the subjects and target market are similar in terms of the relationship under study. Such is the situation with a convenience sample comprised of a ladies' church group participating in an experiment involving certain products of interest to housewives.

Convenience samples are not a major issue in industry or practitioner research. The reasons are: (1) They are used relatively infrequently, and (2) when they are used, they are typically very similar to the marketplace population of interest. In contrast, convenience samples are often a major topic of concern in academic research. The reasons are: (1) convenience samples for academic research are often comprised of college students,[3] and (2) as a result, their demographics and psychographics are usually quite dissimilar to the marketplace population at large.

Student Samples. While college students are not the ideal subjects in terms of external validity, they are often very good subjects. It depends on the situation.[4] Situations where they are an acceptable sample include:

1. When findings serve as a foundation or pioneering effort to be tested with other types of subjects.
2. When products and behaviors of interest to the researcher are characteristic of such subjects.
3. When the research addresses fundamental marketing questions such as cognitive information processing related to marketing stimuli.

The earlier Wheaties example is a situation where college students are an acceptable convenience sample. The findings are likely to be tested with other types of subjects, the product is characteristic of the subjects, and the research addresses a fundamental marketing question. Specifically, the dependent variable in the Wheaties experiment regards measures of eye movement in response to package design characteristics. This dependent variable is likely to be a somewhat universal factor that permits generalization from a convenience sample comprised of college students to the consumer population in general. Furthermore, while a student sample is frequently not justified from a strict research methodology perspective, there is a reality consideration in that student subjects are less expensive in terms of time and money.

There are often concerns about college students as subjects because of the type of research and its inherent goals as opposed to the type of subjects. The type of research that most often uses college students as subjects is academic research. As compared to experiments conducted by practitioners (which usually is applied research), a goal of academic research (which usually is basic research) is to add basic knowledge to the marketing discipline. The results are therefore intended to be generalizable to a much larger spectrum of the marketplace over a much longer period of time. These generalizations therefore involve fundamental consumer behaviors. As such, they are more difficult to identify than the typical research of practitioners, which involves more narrow, short-term behaviors specific to their particular product.

There are some situations in which the use of college students as subjects is readily inappropriate and can never be justified. These situations include when the findings are intended to be generalized to specific marketplace

behaviors characteristic of people who are very dissimilar to the student subjects.[5] An example is to use students in an experiment to generalize about decision-making styles of established business executives in response to marketing efforts, such as presentation styles of professional sales personnel. In such experiments, students have no related experience that is characteristic of the population of interest.

NOTES

1. See earlier, well-accepted material presented by Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1966), pp. 5–6.
2. At this stage of research, the best advertisement in terms of quality can really only be a "rough draft." Even when a very good rough draft (i.e., rough cut) of an advertisement is used, there is a chicken–egg problem. It may lack the major star power (such as an athlete or actor) and expensive production enhancements that will be obtained after a decision is made about which of these three advertisement appeals is most effective.
3. William H. Cunningham, Thomas W. Anderson, Jr., and John H. Murphy, "Are Students Real People?" *Journal of Business* (July 1974), pp. 399–409; Ben M. Enis, Keith K. Cox, and James E. Stafford, "Students as Subjects in Consumer Behavior Experiments," *Journal of Marketing Research* (February 1972), pp. 72–74.
4. Charles W. Lamb, Jr., and Donald E. Stem, Jr., "An Evaluation of Students as Surrogates in Marketing Studies," paper presented at the Association for Consumer Research Conference, 1980; Donald E. Vinson and William J. Lundstrom, "The Use of Students as Experimental Subjects in Marketing Research," *Journal of the Academy of Marketing Science* (Winter 1978), pp. 114–125.
5. Fred W. Morgan, Jr., "Students in Marketing Research: Surrogates vs. Role-Players," *Journal of the Academy of Marketing Science* (Summer 1979), pp. 255–264.

Chapter 5

Experimental Designs and Settings Category

Many types of experiments are conducted in marketing research. An important difference between them often is the level of control, which in turn, usually corresponds to the level of complexity. But the goal of research is always to conduct the proper experiment with the proper level of control. Furthermore, the type of experiment conducted for any particular marketing research project is influenced by the objectives and questions posed in the problem definition stage of the marketing research process, the company's financial budget for marketing research, and the time constraints for when the decision makers require this particular research information to assist in their decisions.

In light of these considerations, the types of experiments available for marketing research are classified in this book according to four categories: setting, independent variables, dependent variables, and subject assignments. These categories are not mutually exclusive, since experiments within different categories generally can be combined and varied to achieve the proper control for a particular research project. Therefore, a major aspect of experiments is the experimental design.

EXPERIMENTAL DESIGNS

Conducting an experiment offers an array of alternatives. This array is large for two reasons: (1) Each type of experiment has variations and (2) types of experiments in different categories are combined. Part of conducting an experiment is the necessity to plan and select the combination of types and variations of types that is most appropriate for the situation at hand. This planning and selecting is commonly referred to as designing the experiment. Designing the experiment is the activity involved in planning and preparing an experimental design. In turn, the experimental design is the comprehensive specifications to conduct an experiment.[1]

Designing the Design

Do not be confused by the verb "designing" (as in planning the experiment) and the noun "design" (as in the plan prepared for conducting the experiment). The distinction between these two concepts is more than grammatical. Designing the experiment represents a process of planning activity. This process involves determining the most appropriate setting and components, constructing a "blueprint" for conducting the experiment, and modifying the plans as necessary. Limitations of time and money are common resource constraints, so some flexibility is necessary to make trade-offs between the ideal and the doable.

The experimental design represents the final plan or blueprint prepared for the entire experiment. It is an important document that should be a comprehensive plan of the experiment and the activities necessary to conduct the experiment. It should be thorough, detailed, foolproof, realistic, capable of modifications as necessary, and, above all, able to provide data appropriate for the defined problem.

The reference above to an experimental design being a blueprint is because an architect's blueprint specifies the building to be built, its location, materials, and construction procedures. Similarly, the experimental design specifies the experiment to be conducted, its setting, components, and procedures to perform the experiment. In both situations, common resource constraints are limited time and money, so some flexibility is necessary to make trade-offs between the ideal experiment and that which is feasible.

An experimental design represents a "blueprint" (i.e., plan) that identifies items and activities necessary to conduct the entire experiment. Its content usually includes specifics about:

- setting—laboratory versus field
- components—independent variables (values and manipulations) and dependent variables (type and measurements)
- subjects—type, and process of assignment to experimental treatments
- other pertinent factors—such as identification and control of potential extraneous variables
- procedures—to conduct the entire experiment
- other details as necessary

Diagramming the Design

Major components and procedures of an experiment are often illustrated with standard diagram techniques.[2] These diagrams use letters of

the alphabet to represent experimental treatments (i.e., manipulation of the independent variables), dependent variable measures, and subject assignments:

X—Indicates exposure of a group of subjects to an experimental treatment.

O—Indicates an observation (i.e., a measurement) of the dependent variable.

R—Indicates random assignment of subjects.

Passage of time is indicated by the order or sequence of these letters along a horizontal line. A letter located first on a horizontal line represents an earlier time period than a letter located second or third. When two or more horizontal lines are presented in a diagram, the letters vertical to each other represent concurrent occurrence. The mechanics of these diagrams will be clear later in this chapter when different types of experimental designs are diagrammed and discussed.

Experimental design in many ways parallels research design within the marketing research process. Both represent planning stages which, when completed, begin performance of the respective research projects. Specifically, to conduct an experiment means to perform a myriad of activities categorized into the following four stages:

1. Formulate hypotheses (to be tested)
2. Design the experiment
3. Perform the experiment
   3.1 Collect the Data
   3.2 Analyze the Data
4. Present a report (of the experiment)

SETTINGS CATEGORY

Two types of settings identify experiments: laboratory and field. Laboratory experiments align with a highly controlled setting (environment) that is important for high internal validity. Field experiments align with a realistic or natural setting that is important for high external validity. The ideal experiment, that permits both a completely natural setting and total control to eliminate alternative explanation of the results, does not exist. However, while Exhibit 5.1 summarizes extremes, trade-offs can be made to increase reality in laboratory experiments and to increase control in field experiments.

Exhibit 5.1
Laboratory Experiments versus Field Experiments

```
Left column is                          Right column is
characteristic of                       characteristic of
FIELD EXPERIMENTS                       LABORATORY EXPERIMENTS

                   CONTINUUM RANGING FROM:

Realistic,                              Artificial,
natural environment.....-to-......unnatural environment

Low control of                          High control of
extraneous                              extraneous
variables..............-to-.......variables

Lower internal                          Higher internal
validity...............-to-.......validity

Higher external                         Lower external
validity...............-to-.......validity
```

Laboratory Experiments

A laboratory experiment is an experiment conducted in a setting expressly prepared for the research project. Its main advantage is research control. It is theoretically possible to achieve total control of effects that are due to extraneous variables. In reality, control that is achieved is not total but it is relatively thorough compared to experiments conducted outside a laboratory.

The main disadvantage with this type of experiment is artificiality. As control is increased, the environment in which data are collected becomes more artificial. One consequent is that subjects are almost always aware that they are participating in a research study. As a result, the collected data may not accurately represent realities of the marketplace. Efforts are made to increase the realism of laboratory experiments in an attempt to counter artificiality. Despite these efforts, laboratory experiments generally remain less realistic then field experiments.

Laboratory experiments are used to study all types of marketing questions. One exception is questions pertaining to formal test marketing, which by definition involves testing a product in an actual marketplace. Uses in marketing research are nearly limitless. A few examples include:

1. A laboratory experiment with Bounty paper towels. One independent variable (package design) with two values: (a) package stating "50% MORE ABSORBENCY THAN COMPETING BRANDS" and (b) package stating "MORE DECORATIVE COLORS THAN COMPETING BRANDS." Measures of dependent variable: Attitudes and intent to purchase Bounty paper towels. Procedures: Shoppers intercepted at a mall, invited into a back room, and shown one of the two experimental conditions.
2. A laboratory experiment with a potential new television series for the CBS television network. One independent variable (title name) with two values: (a) "STEEL COLLAR MAN" and (b) "DX-5 STEEL COLLAR MAN." Measures of dependent variable: Liking and intent to watch the program, assessed via electrical-mechanical push buttons and paper-and-pencil questionnaires. Procedures: Tourists from across the United States, while visiting New York City or Los Angeles, recruited to a special showing at the respective N.Y.C. or L.A. studio headquarters. Subjects are shown the television show with one of the two titles/experimental conditions.
3. A laboratory experiment for the Young and Rubicam advertising agency in New York City. One independent variable (print advertisement for a new style of home telephone to be marketed by the General Electric Company) with four values: four different versions of the telephone. Measures of dependent variable: Interest in advertising, assessed with electrodes attached to the subjects' heads to record brain waves and with special cameras to record eye movements while the subjects' heads are held perfectly still in a special brace. Procedures: A local PTA club is invited to view advertisements for a donation to the club from the company.
4. A laboratory experiment for the J. Walter Thompson advertising agency in Chicago. One independent variable (television advertisement for a new General Motors car) with two values: two different versions. Measures of dependent variable: Liking and disliking for the commercial and product, assessed while watching commercials and operating a switch attached to subjects' chairs. Procedures: College students from a local university are invited to a special theater to view commercials.

Field Experiments

A field experiment is an experiment conducted in as natural a setting as possible. Once a location is selected, actions are expressly made so its naturalness is not changed any more than necessary to still be able to conduct the experiment. An advantage is that subjects often do not know they are involved in a research project when they participate in a field experiment. The collected data are therefore more likely to represent realities of the marketplace. Furthermore, field experiments are especially applicable for test marketing, since by definition test marketing involves testing a product in an actual marketplace. As a result, settings for field experiments often involve actual cities, as presented in Exhibit 5.2.

A disadvantage of field experiments is that control of extraneous varia-

Exhibit 5.2
Cities Most Frequently Used for Test Marketing

Akron, OH	Marion, IN
Ann Arbor, MI	Melbourne, FL
Asheville, NC	Midland, TX
Austin, TX	Mobile, AL
Bangor, ME	Montgomery, AL
Beaumont, TX	New Orleans, LA
Boise, ID	Oklahoma City, OK
Buffalo, NY	Orlando, FL
Cedar Rapids, IA	Philadelphia, PA
Charleston, WV	Pittsfield, MA
Chicago, IL	Portland, OR
Colorado Springs, CO	Providence, RI
Columbus, OH	Raleigh, NC
Dallas, TX	Richmond, VA
Decatur, IL	Rockford, IL
Detroit, MI	St. Louis, MO
Durham, NC	Salem, NC
Elkhart, IN	Salt Lake City, UT
Evansville, IN	San Francisco, CA
Fort Collins, CO	Scranton, PA
Fort Wayne, IN	Sioux Falls, SD
Grand Junction, CO	Spokane, WA
Greensboro, NC	Springfield, IL
Hartford, CT	Syracuse, NY
Huntsville, AL	Tampa, FL
Jacksonville, FL	Troy, NY
Kansas City, MO	Washington, DC
Las Vegas, NV	Wichita, KS
Little Rock, AK	Yakima, WA
Lubbock, TX	

Source: "The Nation's Most Popular Test Markets," Sales & Marketing Management, March 1989, pp. 65-66.

bles is almost always substantially less than with laboratory experiments. It is not possible to control all the extraneous variables in a natural marketing environment, since the variables are too numerous and too complex. Therefore, a trade-off exists between increases of external validity and decreases of internal validity.

Field experiments are used to study all types of marketing questions. A few examples include the following:

1. A field experiment with Jif Peanut Butter. One independent variable (label information) with two values: (a) label stating "LESS SUGAR" and (b) label stating "MORE FLAVOR." Measures of dependent variable: Product sales and, later, an attitude questionnaire of shoppers. Procedures: Jars with the two different labels placed side by side in the same grocery store (such as in the Alpha Beta chain in Los Angeles or the Kroeger chain in Atlanta). Two alternative conditions: (a) jars with different labels placed in different stores in the same city, and (b) jars with different labels placed in different cities.

2. A field experiment with Hermann Joseph's beer by the Coors Beer Company. One independent variable (the product) with one value: the product. Measures of dependent variable: Product sales, measured/monitored by location and store type. Procedures: Test marketing project for new product under consideration at Coors Beer Company. Locations were Albuquerque, Phoenix, Austin, and Seattle. Marketing efforts were kept as constant as possible in all the test cities. Because of the reality of this field experiment, the company used these sales figures to generalize national sales in the event they decided to commercialize the product and roll it on a national scale.
3. A field experiment with a potential new television sitcom series for the NBC television network. One independent variable (program content) with two values: (a) topic on death and (b) topic on marriage. Measure of dependent variable: Liking of program and intent to watch, assessed via telephone calls to the viewers' homes. Procedures: Using "split-signal" cable television technology to deliver/broadcast one of two episodes of a potential television series to only particular homes (on a block-by-block basis and a city-by-city basis). Afterwards, the households receiving this television program are called.
4. A field experiment by Young and Rubicam advertising agency for a new style telephone to be marketed by General Electric. One independent variable (television commercial) with three values: three different versions of the product. Measures of dependent variable: Actual sales of the telephone. Procedures: Using the same "split-signal" cable television as indicated above, each of three different television commercials are delivered to homes in different communities. Each commercial requests immediate purchase action, accompanied by an 800 telephone number to place orders. By using both cable television and telephone ordering systems, this field experiment provides strong measures of each commercial's effectiveness.

Managerial Implication

Experiments are conducted in both laboratory settings and field settings. In fact, both types of experiments may be conducted in conjunction with each other. Often a laboratory experiment will provide findings that are investigated later by a field experiment.[3] One such example is presented in Exhibit 5.3. It should be noted that variations in experimental designs yield somewhat hybrid types of experiments. As a result, concerns about external validity associated with laboratory experiments and internal validity associated with field experiments can be moderated when experiments are properly planned and designed.

Exhibit 5.3
Use of Laboratory Experiments and Field Experiments: A Firsthand Experience Example

In 1990 the author of this textbook was working for a marketing research firm with operations in the United States, Europe, and Japan. Universal Pictures was at that time about to distribute a movie titled DAD. A Steven Spielberg production, it starred Jack Lemmon as the father and Ted Danson as the son, and was about to be marketed throughout the United States and Europe.

Before marketing the movie to the public, the company commissioned the above mentioned marketing research firm to conduct research to determine which of two promotional appeals would work best. To be confident of its decision, both a laboratory experiment and a field experiment were conducted for this research question. These experiments were as identical as possible. The independent variable (commercial headline of a 30-second television advertisement) had two values with only minor differences: (1) an off-camera voice stating: "...SOMETIMES THE GREATEST MAN YOU EVER MEET IS THE FIRST ONE" and (2) an identical advertisement except for the off-camera voice stating: "...THE GREATEST MAN JOHN TREMONT [star of the movie] EVER MET WAS THE FIRST ONE."

Both the laboratory experiment and the field experiment were conducted in the United States and Europe. The laboratory experiment involved showing video tapes to people in malls in eight cities in the United States and in shopping areas in eight cities in western Europe. Subjects were afterwards asked to complete questionnaires that inquired about interest in seeing the movie. The field experiment involved actual broadcasting of the two different advertisements in three different markets in the United States and in three different markets in western Europe. Afterwards, telephone calls were made to households to assess interest in seeing the movie.

Note: While the actual data are proprietary, marketing executives decided to use the first headline in Europe ("...SOMETIMES THE GREATEST MAN YOU EVER MEET IS THE FIRST ONE") and the second in the United States ("...THE GREATEST MAN JOHN TREMONT EVER MET WAS THE FIRST ONE)."

NOTES

1. Closely related to experimental design is the phrase "designing the experiment," which is the activity involved in planning and preparing an experimental design. Do not be confused by the verb "designing" (as in planning the experiment) and the noun "design" (as in the plan prepared for conducting the experiment). In other words, designing the experiment is a process of planning activity and the experimental design is the end product of that planning.
2. Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1966), pp. 13–25.

3. For a more generic and more thorough discussion of the respective information provided by laboratory and field experiments see Alan G. Sawyer, Parket M. Worthing, and Paul E. Sendark, "The Role of Laboratory Experiments to Test Marketing Strategies," *Journal of Marketing* (Summer 1979), pp. 60–67.

Chapter 6

Independent Variable Category

Of the major components, the independent variable(s) can be considered the first to be selected or decided upon when designing an experiment. Decisions about independent variables are in direct response to the hypothesis to be tested. Given the key position held by independent variables in an experiment, it is reasonable to consider this the "Principle Category" of experiments. It also is reasonable to consider this the principle category of experiments because these experiments are by and large the standard against which experiments are compared and altered.

Two types of experiments comprise this category. The most fundamental is an experiment comprised of one independent variable known as a "Pretreatment–Posttreatment Control Group Experiment." In contrast, a variation of this fundamental experimental design, which almost always represents greater complexity, is the "factorial experiment." A primary difference is that the factorial experiment is comprised of more than one independent variable.

This chapter also is designated the principle category because it delineates the minimum components and procedures that qualify a research activity as an actual experiment. This delineation is necessary because the term "experiment" has lost its true meaning among many people. To these people, experiment commonly means "to simply try out an idea," with no regard for the myriad considerations and procedures required to conduct an actual experiment.

PRETREATMENT–POSTTREATMENT CONTROL GROUP EXPERIMENT

Pretreatment–Posttreatment Control Group Experiment has an experimental group and a control group, and the dependent variable is measured twice for all subjects. An advantage of this type of experiment is that it meets the requirements to infer causality at minimal expense and time. A disadvantage is the limited number of variables and values that are investigated in a particular project. This type of experiment is often designated in terms of "pretreatment/before" and "posttreatment/after," which means that Pretreatment–Posttreatment Control Group Experiment and Before–After Control Group Experiment are synonymous.

The Pretreatment–Posttreatment Control Group Experiment is the classic experimental design. It is time-honored and exemplary as a model against which other experimental designs can be compared. Exhibit 6.1 lists the minimum components necessary to conduct the most basic version of this experiment, which has one independent variable and two experimental conditions. The independent variable only has one value. The two experimental conditions are the treatment group and control group, to which subjects are randomly assigned.

Exhibit 6.2 diagrams the minimum procedure to conduct an experiment, beginning with random assignment of subjects to either the experimental group or control group (indicated by "R"). The dependent variable is measured for the experimental group (indicated by "O_1") and control group (indicated by "O_3") at the same time (indicated by the vertical arrangement of "O_1" and "O_3").

Subjects in the experimental condition then receive the experimental treatment (indicated by "X") and subjects in the control condition receive no experimental treatment (indicated by the blank space vertical to the experimental treatment "X"). Exhibit 6.2 also illustrates that, after an equal amount of time elapses for both groups (indicated by the equal space between the first and second measures for each group), the dependent variable is measured a second time for the experimental group (indicated by "O_2") and the control group (indicated by "O_4"). These measures are assessed at the same time (indicated by the vertical arrangement of "O_2" and "O_4".

For illustration, consider an experiment for Heinz ketchup. It has one independent variable (package size) with two values: the current 8-ounce size and a potential new 20-ounce size. The procedure exposes the experimental treatment group of subjects to the 20-ounce package, while the control group is exposed only to the status quo size of 8 ounces. Dependent variable is intent to purchase Heinz ketchup, measured by questionnaire, before and after exposure to the 20-ounce package (experimental group),

Exhibit 6.1
Minimum Components Necessary to Conduct a Basic, Classic Experiment

```
The following five components are the minimum required
for the procedure to conduct an experiment
(specifically, the most basic version of the classic
experimental design known as the Pretreatment-
Posttreatment Control Group Experiment):

    (1)  One experimental group

    (2)  One control group

    (3)  Randomized assignment of subjects to groups

    (4)  Pretreatment measure of experimental and
         control groups

    (5)  Posttreatment measure of experimental and
         control groups
```

and at corresponding times with no change to the current 8-ounce package for the control group.

Calculating Effect

Effect of the experimental treatment for a Pretreatment–Posttreatment Control Group Experiment is revealed by the difference between the first and second measures for each of the two groups. It equals: $(O_2 - O_1) - (O_4 - O_3)$.

In the Heinz ketchup example, assume that before exposure to the experimental treatment, 10% of all subjects intend to buy the product (O_1 = 10% and O_3 = 10%). After the experimental treatment 20% intend to buy Heinz ketchup (O_2 = 20%). After an equal amount of elapsed time, 12% (O_4) of the control group intend to purchase the product. To determine the experimental treatment effect, the difference between the first and second measures is calculated for each of the two groups:

$$\begin{aligned}\text{experimental treatment effect} &= (O_2 - O_1) - (O_4 - O_3)\\ &= (20\% - 10\%) - (12\% - 10\%)\\ &= 8\%\end{aligned}$$

These calculations indicate that the experimental treatment (a larger, 20-ounce package) caused an 8 percentage point increase in the dependent variable: "stated intent to buy Heinz ketchup."

Exhibit 6.2
Diagram of the Minimum Components and Procedure Necessary to Conduct a Basic, Classic Experiment

```
        ┌──> (Experimental Group)   O1   X   O2
R:  >───┤
        └──> (Control Group)        O3       O4
____________________________________________

Using the alphabet letters and configuration explained in
the earlier discussion about "Experimental Design":

    "R" indicates random assignment of subjects.
    "O" indicates measurement of the dependent variable.
    "X" indicates exposure of subjects to the
        experimental treatment.
____________________________________________

Note: This diagram represents the procedure necessary to
conduct an experiment; specifically, the most basic
version of the classic experimental design known as the
Pretreatment-Posttreatment Control Group Experiment.
```

Importance of the Control Group

The increase of two percentage points among the control group subjects (O_3 = 10% and O_4 = 12%) in the Heinz ketchup experiment highlights the value of the classic experiment. Since the control group was administered no experimental treatment, the observed increase likely was caused by one or more extraneous variables—statistical regression, selection, instrumentation, history, maturation, and/or mortality—that likely caused the same effect among experimental group subjects. However, the effect of these extraneous variables does not necessarily pose a problem because:

1. Random assignment distributes the effect of these variables equally among the two groups.
2. Total extraneous variable effect is indicated and reconciled when effect on the dependent variable is calculated by subtracting the change among the control group from the change among the experimental treatment group: $(O_2 - O_1) - (O_4 - O_3)$.

Managerial Implication

This type of experiment (Pretreatment–Posttreatment Control Group) is powerful, but overreliance on it can result in wrong decisions among unenlightened decision makers. The reason is that the extraneous variable associated with testing effect, specifically reactive/interactive testing effect, is not controlled because the experimental and control groups are exposed to the initial dependent variable measure in the same way and at the same time.

The impact of the initial measure, on issues such as increased awareness or sensitivity, cannot be determined by this experimental design. Therefore, reality in the marketplace might differ, since exposing subjects to the initial questionnaire may trigger "selective perception" not shared by people in the marketplace. As discussed in consumer behavior courses, selective perception occurs when people notice events and stimuli more for which an interest exists. In this situation, exposure to the initial measure of the dependent variable (O_1 and O_3) could cause an interest to exist. This interest then might cause subjects to respond differently to the second measure (O_2 and O_4), about intent to purchase Heinz ketchup, than consumers in the marketplace who are not similarly exposed to an initial questionnaire and thus not similarly sensitized or aware.

Heinz Ketchup Experiment

Consider again the Heinz ketchup experiment. Results show a larger percent of subjects "intend to buy the product" when marketed in the larger, 20-ounce package than the current 8-ounce package size. However, a decision to repackage the product in the larger size and to discontinue the current small size might be incorrect, because it requires two risky assumptions: (1) conditions in this experiment are for all practical purposes the same as conditions in the marketplace, and (2) expressed "intent to buy" will translate into actual purchase behavior, with an eight percentage point difference in the market purchasing Heinz ketchup when sold in the 20-ounce package.

Further Implications

Decision makers should be aware of implicit assumptions and their impact on related decisions, testing effects, and trade-offs necessary to eliminate them. For example, if testing effects are judged to reduce internal validity significantly, another type of experiment can be conducted to control them. An alternative is the Solomon Four Group Experiment discussed later in this chapter, but the trade-off is increased complexity and costs.

Decision makers also should be aware that an important variation of this basic, classic experiment is to increase the number of independent variable values. This variation represents better the reality of a marketing decision maker. For example, the Heinz ketchup example could involve multiple potential values of interest to a marketing manager: 2-ounce size, 6 ounces, 10 ounces, 14 ounces, 24 ounces, and so on. Another variation is to increase the number of independent variables, and conduct a factorial experiment that might include values/levels for package color, shape, and durability, price level, promotion appeal, distribution outlets, and so on, as well as package size.

The reason for these variations is that the fundamental experiment involves one independent variable. However, more complex variations that increase the number of values of the independent variable are often necessary to better correspond with the reality of marketing management, who often require that an experiment be designed to study more than one independent variable (e.g., package size, package color, package strength, price level, promotion appeal, etc.). In fact, the number of questions that marketing decision makers often, at least initially, pose in the problem definition stage are almost limitless. But the corresponding independent variables in any one experiment are, for practical purposes, generally limited to a maximum of four or five. The reason for this general maximum is that as an experiment increases in its number of independent variable values and/or independent variables, the number of necessary subjects and the difficulty of analyzing and interpreting the data increase geometrically.

The type of experiment that goes beyond the fundamental experiment, to involve more than one independent variable, is referred to as a "factorial" experiment. It is the type of experiment most frequently used in marketing research. The reason for its frequency of use is that it corresponds relatively well with the reality of multiple marketing questions and thus multiple independent variables, while at the same time maintaining the minimum components and procedure necessary to conduct a basic, classic experiment.

FACTORIAL EXPERIMENT

A factorial experiment involves simultaneous study of more than one independent variable. It also permits simultaneous study of more than one value for each independent variable. Two related terms are factor—an independent variable in a factorial experiment, and level—a value of an independent variable in a factorial experiment. Factorial experiments are the most frequently used because they accommodate multiple variables (factors) reasonably well that are relatively characteristic of marketing decisions.

To illustrate, consider that Macintosh has decided to market a new notebook computer with production cost of $1,000 per unit. In designing the marketing plan, company executives question the effect on "intent of consumers to purchase the computer" (i.e., dependent variable) that will be caused by the type of distribution outlet that sells the computer (i.e., independent variable or factor) and the type of carrying case provided with the computer (i.e., independent variable or factor). Assume:

- Two alternatives (i.e., values or levels) for the distribution factor: discount stores such as K-Mart and Target, and specialty computer stores such as ComputerLand and BusinessLand.
- Three alternatives (i.e., values or levels) for the carrying case factor: hand strap model, shoulder strap model, and hand-and-shoulder strap model.

This research project is a 2 × 3 (read "two by three") factorial experiment. The numbers indicate two factors (independent variables) with the first factor comprised of two levels (values) and the second factor comprised of three levels.

A strength of a factorial experiment is that it provides dependent variable data for every possible combination of independent variables (i.e., factors) and their values (i.e., levels).[1] These combinations are illustrated in the prototype table presented in Exhibit 6.3. It is helpful to arrange the factors and their levels by rows and columns. The result is that each intersection of a row and a column identifies one of the possible combinations. These intersections or combinations can be further identified by using "A" and "B" designations, which identify respective cells with precise coordinates of rows and columns, such as A_1B_1, A_1B_2, and so on.

In most all factorial experiments, different groups of subjects are exposed to different treatment combinations. To calculate the number of groups required, the factors and levels are simply multiplied. In this example, a 2 × 3 factorial experiment would require six experimental groups; as shown by the number of cells indicated in Exhibit 6.3. In each of the six cells (i.e., by the number of cells indicated in Exhibit 6.3. In each of the six cells (i.e., six experimental groups), measures of the dependent variable caused by manipulation of the independent variables are recorded for the different treatment levels (i.e., combinations of independent variable values).

Combination of Variables

Exhibit 6.3 suggests the ability of factorial experiments to provide dependent variable data for all combinations of factors and levels under study.[2] The cells in this table show the percentages of subjects who indicate affirmatively their intent to purchase the notebook computer when it is sold, at two different types of stores with three different types of carrying case. Each cell, except in experiments with repeated assignment, represents measures of the dependent variable from a different group of subjects. Therefore, the number of groups required for an experiment is the number of factors and levels multiplied, which is six groups for a 2 × 3 factorial experiment.

Effects: Main and Interaction

Factorial experiments permit investigation of main effects and interaction effects. A main effect is the impact on the dependent variable caused by an

Exhibit 6.3
Prototype Table Illustrating a 2 × 3 Factorial Experiment for a Notebook Computer

Table # [actual number to be provided by researcher]

Intent to Purchase Notebook Computer by Type of Distribution Outlet (Store) and by Type of Carrying Case

	Carrying Case [Factor "B"]			
Store [Factor "A"]	Hand Strap	Shoulder Strap	Hand and Shoulder	overall mean percent
Discount Stores	40%	20%	25%	28.3
Specialty Stores	10%	30%	25%	21.6
overall mean percent	25.0	25.0	25.0	

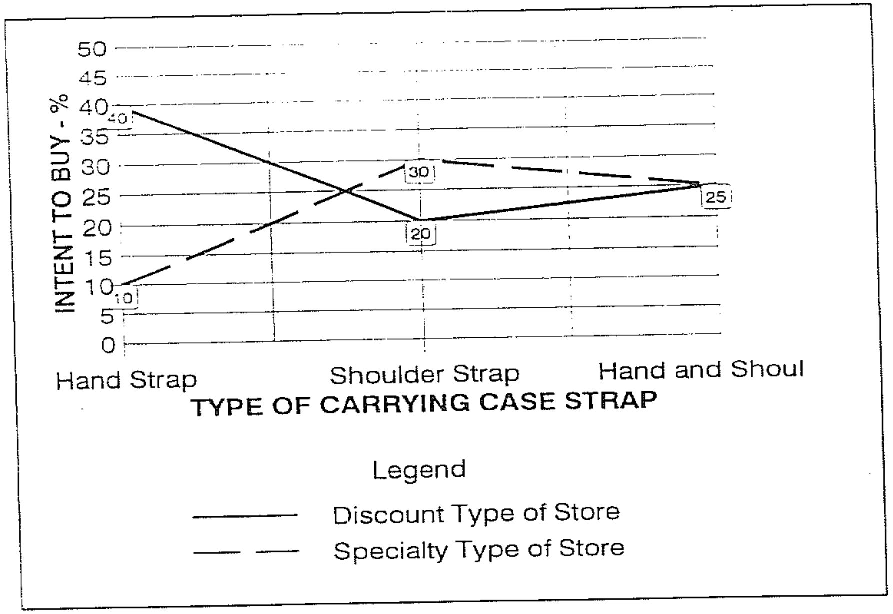
Note:

Each cell represents measures of the dependent variable, i.e., percentage of subjects indicating, affirmatively, their "Intent to Purchase the Notebook Computer" under study.

independent variable. Exhibit 6.3 illustrates a main effect for factor "A" (type of store) but not for factor "B" (type of carrying case). Specifically, a larger percent of subjects state they intend to buy the computer if sold through discount stores (28.3%) than if sold through specialty stores (21.6%). But the percent is the same if the carrying case has a hand strap, shoulder strap, or hand and shoulder strap (25.0%).

Factorial experiments have the ability to go beyond main effects to identify interaction effects. An interaction effect is when the impact of an independent variable differs with different combinations of independent variables and/or their values. For example, the type of store and type of carrying case may each have a separate impact on sales of a notebook computer, which together is different.

Exhibit 6.4
Line Graph of Data Collected from a 2 × 3 Factorial Experiment



The data in Exhibit 6.3 also illustrate interaction effects. While the overall mean percent is greater for discount stores (28.3%) than specialty stores (21.6%), an interaction effect reveals the opposite with even greater percentages. If the computer is sold with a shoulder strap carrying case, a larger percent of subjects express that they intend to buy the computer if it is sold at specialty stores (30%) than if it is sold at discount stores (20%).

Although total intent to purchase is the same regardless of carrying case type (25.0%), interaction effects reveal differences. First, twice the percent of "discount store subjects" express that they intend to buy the computer if it is sold with a hand strap carrying case (40%) than if sold with a shoulder strap carrying case (20%). Second, this relationship is reversed for the "specialty store subjects," in which an even larger percent express an intent to buy the computer if sold with a shoulder strap carrying case (30%) than if sold with a hand strap case (10%). Third, four times as many subjects express intent to purchase the computer with a hand strap case if sold at discount stores (40%) than if sold at specialty stores (10%).

Visual presentations help interpret data from factorial experiments. For example, the Exhibit 6.4 line graph clearly shows the effect on sales (from the Exhibit 6.3 table data) for each value of the two independent variables. Such graphs also indicate interaction effects when lines cross as in Exhibit 6.4 or, less dramatically, differ in angle or direction.

Managerial Implication

The factorial experiment is the "workhorse" of marketing research experiments, and is applicable to many of the other types of experiments discussed in this chapter. It is an effective and efficient method to provide information to assist marketing decision makers. Its effectiveness involves the ability to investigate multiple independent variables, each with multiple values, simultaneously in one study. Compared to conducting a series of individual experiments with a single independent variable, time and money are saved. Furthermore, powerful information about interaction effects, not possible with any series of individual experiments, is provided.

The adage that the "whole is greater than the sum of the parts" applies to both the efficiency and effectiveness of the factorial experiment. Especially valuable is the aspect of effectiveness regarding simultaneous investigation of multiple independent variables and multiple values. The result is powerful information about interaction effects, which would not be possible with a series of individual experiments. For example, in the above computer notebook example,[3] the Macintosh computer company has the opportunity to substantially increase its marketing success by marketing one type of carrying case (the hand strap) through discount stores and another type (the shoulder strap) through specialty stores. But without information about interaction effects, the type of carrying case appears unimportant.

A disadvantage of factorial experiments is their complexity if many variables are included. This complexity translates to greater expense in time and money to conduct the experiment, and more difficulty to interpret the data.

NOTES

1. On select occasions, and with superior command of the methodology, it has been shown that instead of every possible combination, a subset can be satisfactorily utilized to simplify the experiment. See Charles W. Holland and David W. Cravens, "Fractional Factorial Experimental Designs in Marketing Research," *Journal of Marketing Research* 10 (August 1973), pp. 270–276.
2. Ibid.
3. For an example, with a more technical presentation of the value specifically provided by a factorial experiment see J. B. Wilkinson, J. Barry Mason, and Christie H. Paksoy, "Assessing the Impact of Short-Term Supermarket Strategy Variables," *Journal of Marketing Research* 19 (February 1982), pp. 72–86.

Chapter 7

Dependent Variable Category

Dependent variables serve as a reasonable way to identify a type of experiment that focuses on the period of time that the dependent variable is measured: either before and after the experimental treatment is administered, or only after. Generally, experiments that involve before-and-after measures represent a more complex type of experiment than those that involve only-after measures.

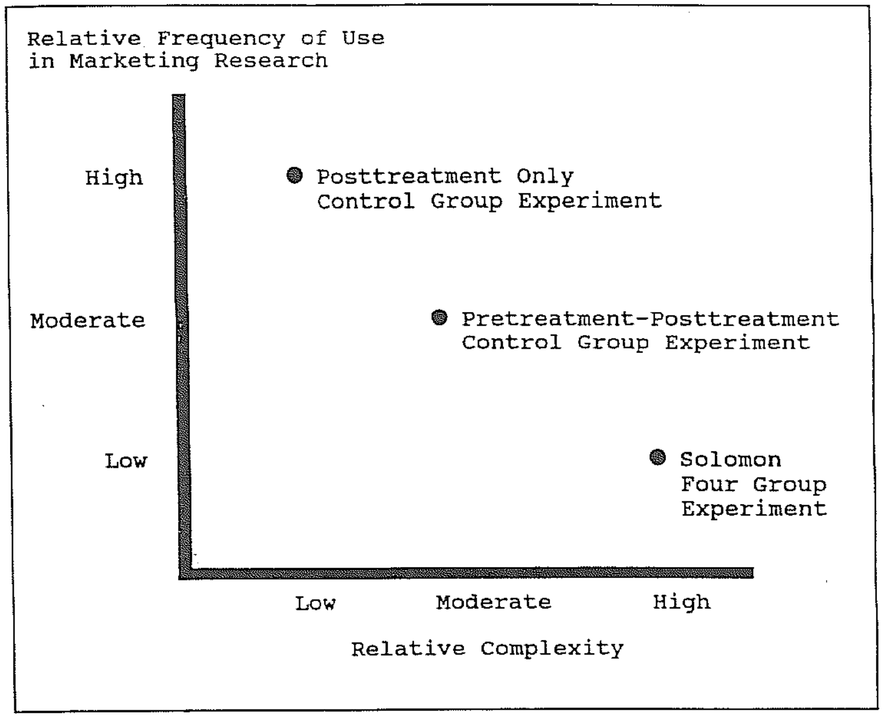
Marketing research that utilizes this type of experiment most frequently uses a type known as a "Posttreatment Only Control Group Experiment." Its primary variation from the basic, classic experimental design (discussed earlier as the Pretreatment–Posttreatment Control Group Experiment within the principle category of experiments) is that a pretreatment measure of the dependent variable is not employed.

Another type of experiment in this category, which represents even greater complexity and less use in marketing research, is the "Solomon Four Group" experiment. The reason it is more complex than the basic, classic design is because it has two control groups and two experimental groups. To illustrate the relative complexity and use of these experiments, Exhibit 7.1 lists these two types of experiments in decreasing order for their relative frequency of use in marketing research and increasing order for their relative complexity, with the Pretreatment–Posttreatment Control Group Experiment in between the two.

POSTTREATMENT ONLY CONTROL GROUP EXPERIMENT

The Posttreatment Only Control Group Experiment has an experimental group and a control group, and measures the dependent variable only once

Exhibit 7.1
Relative Frequency of Use, and Relative Complexity of Three Types of Experiments Conducted in Marketing Research



for all subjects. It is the most frequently used type of experiment in this category in marketing research and may be, at times, the only practical experiment to conduct. Its primary difference from the classical experimental design (Exhibit 6.1 and Exhibit 6.2) is that a pretreatment measure of the dependent variable is not employed. The word "posttreatment" is often designated "posttreatment" or "after," such that Posttreatment Only Control Group Experiment and After Only Control Group Experiment are synonymous.

Exhibit 7.2 illustrates that all subjects are assigned randomly to either the experimental group or the control group. After an equal amount of time elapses, measures of the dependent variable occur at the same time for both groups (indicated by the vertical arrangement of O_1 and O_2). While the elapsed time and measurement times are the same for both groups, their experiences are different. During this time, the experimental group is exposed to the experimental treatment ("X"), while the control group is not exposed to an experimental treatment (indicated by the corresponding blank space).

Exhibit 7.2
Diagram Illustrating a Posttreatment Only Control Group Experiment

```
        ┌──> (Experimental Group)  X  O1
R: >────┤
        └──> (Control Group)          O2
________________________________________

"R" indicates random assignment of subjects.
"O" indicates measurement of the dependent variable.
"X" indicates exposure of subjects to the experimental
     treatment.
```

Calculating Effect

The effect of the experimental treatment for a Posttreatment Only Control Group Experiment is the difference between the two measures ($O_1 - O_2$). Consider marketing Clearasil acne medication. The product is currently sold in 2-ounce packages, but company executives have approved an experiment to investigate the impact on sales, via purchase intent, that a new package size might cause. The independent variable value (i.e., experimental treatment) is an 8-ounce package size, dependent variable is the percentage of people who plan to buy the product, and the subjects are 100 16-year-old students from the John F. Kennedy High School in Atlanta who are randomly assigned to one of the two groups (as indicated by the "R" in Exhibit 7.2).

After administering the experimental treatment "X" (i.e., after showing the subjects in the experimental group the 8-ounce package size), these subjects are requested to complete a questionnaire ("O_1") to assess their intent to buy the product. At the same time, the control group (i.e., those subjects not shown the 8-ounce package) are requested also to complete a questionnaire ("O_2") to assess their intent to buy the product. Assume that 15% in the experimental group stated "yes" ($O_1 = 15\%$) and 10% in the control group stated "yes" ($O_2 = 10\%$). The effect of the experimental treatment (a larger package size) caused on the dependent variable (intent to purchase) is an increase of 5 percentage points: ($O_1 = 15\%$) − ($O_2 = 10\%$) = 5%.

Managerial Implication

This type of experiment is appropriate when it is necessary to control for testing effects. This necessity occurs when pretreatment measurements are not possible, as in situations when initial measurements would sensitize the subjects to such an extent that the experiment would not be valid.

If the two groups above were given an initial questionnaire about intent to purchase the product, it could have substantially sensitized them to the product and/or brand, and even informed the experimental group that the package size was of interest to the researchers. As a result, the measurements (via the completed questionnaires) that were made could be expected to be different if the subjects had responded to the same questions earlier and were now doing it a second time. Since there were no "initial" or pretreatment measurements, an advantage of the Posttreatment Only Control Group Experiment is that a testing effect cannot occur. This lack of testing effect is a distinct advantage. In fact, a disadvantage of the standard type of experiment (the Pretreatment–Posttreatment Control Group Experiment) is that it does not control for testing effects.

Trade-Offs

A distinct advantage of the Posttreatment Only Control Group Experiment is that no "initial" or pretreatment measures are assessed, so a testing effect cannot occur. The trade-off is that effects caused by other potential extraneous variables during the elapsed time cannot be determined. In comparison, while a disadvantage of the Pretreatment–Posttreatment Control Group Experiment is that it does not control for testing effects, knowledge of other effects is determined through its pretreatment and posttreatment measures.

Regardless, the trade-off is reasonable, because lack of information with the Posttreatment Only Control Group Experiment does not mean lack of control. The control of extraneous variable effects is accomplished through random assignment. Furthermore, with random assignment, effects due to history, maturation, statistical regression, selection, and mortality are all distributed approximately the same for each group of subjects.

Usage, Cost, and Complexity

Avoiding the cost incurred with pretreatment measures contributes to the frequency of use for this type of experiment. For example, pretreatment measures strain the company's research budget and delay the results, possibly beyond the deadline for company executives waiting for information to assist their decision making. Exhibits 7.3 and 7.4 present an actual account of these costs.

For subjects who are already inconvenienced, pretreatment measures can be the "straw that broke the camel's back." They expend effort and time, and usually receive only a complimentary thank you or nominal monetary token. An additional request for pretreatment measures can impact their

Exhibit 7.3
Prohibitive Costs for Pretreatment Measures: A Firsthand Experience Example
(Continued in Exhibit 7.4)

While working in marketing research, the author of this textbook experienced firsthand the sometimes prohibitive costs associated with pretreatment measures. The example presented here is applicable to marketing research projects regardless of industry or product, and involves the movie and television industries.

The following situation is not unusual: The marketing research subjects are people on vacation, recruited to participate in an experiment that requires up to 2 hours of their time. After viewing a movie for about 1 and 1/2 hours, they complete a questionnaire and participate in an abbreviated focus group for about another 1/2 hour or more. The experimental design may specify collecting all data within ten days for 1,000 subjects, to view the movie in groups of 25 or less, in a special viewing room, for which a token souvenir ink pen of nominal value is given.

Time and financial considerations are substantial for the company executives. Therefore, the specified ten days for data collection might be vital to assist various marketing decisions in a timely manner. The financial costs include recruiting personnel, rental of office/viewing space, rental of projection equipment, and huge financial penalties if production of the product gets off schedule.

Time considerations are substantial for the subjects, who give their precious vacation time to watch a movie they know nothing about and may not like. Furthermore, they must expend additional time and effort to complete a paper and pencil questionnaire and discuss their feelings among strangers in a focus group.

data negatively, if they become upset, irritable, or otherwise displeased before completing the experiment.

SOLOMON FOUR GROUP EXPERIMENT

The Solomon Four Group Experiment has one independent variable value, two experimental groups, and two control groups. Exhibit 7.5 diagrams the procedure to conduct this type of experiment, which begins with random assignment of subjects to one of four groups. For the first experimental group and first control group, pretreatment and posttreatment measures are assessed. For the second experimental group and second con-

Exhibit 7.4
Prohibitive Costs for Pretreatment Measures: A Firsthand Experience Example

For the marketing researchers (and indirectly for the decision makers who will use the marketing research), such an experiment typically poses substantial problems. If only one testing room is available, the logistics of collecting data from 1,000 subjects in 10 days in single groups of 25 means that 200 subjects are tested per day. Therefore, at least 8 experimental treatments are conducted (200/25) with a minimum of 2 hours per treatment assuming no delays, which means 16-hour work days at minimum.

Adding any extra time, such as for a pretreatment measure, could easily add 20 extra minutes to each experimental treatment. Then, each experimental treatment becomes 2 hours and 20 minutes. These extra 20 minutes alone, not even considering the extra mental effort to complete the questionnaire, can change the subjects, state of mind during the experimental treatment (of viewing the movie) and during the posttreatment measure (of completing the questionnaires and participating in a focus group).

These extra 20 minutes per experimental treatment also add over 2 hours of work per day for the researchers and their staff (8 treatments times 20 minutes per treatment equals 160 extra minutes). Their workday, just for collecting the data (with absolutely no delays, no equipment breakdowns, and totally smooth moving of subjects in and out of the viewing room) increases from 16 hours to over 18 hours. This extra time in turn incurs additional expenses for personnel (regular and overtime wages) and for facility and equipment rental charges.

In conclusion, the prudence of conducting pretreatment measures from a research perspective must be balanced with the reality of the associated costs.

trol group, only posttreatment measures are assessed. Subsequent calculations reveal experimental effects and extraneous variable effects.

An advantage of this type of experiment is thorough control of extraneous variable effects, in contrast with only a portion controlled by each of the two specific experiments discussed thus far. For example, Pretreatment–Posttreatment Control Group Experiments control effects due to statistical regression, selection, instrumentation, history, maturation, and mortality, but not testing, whereas Posttreatment Only Control Group Experiments control testing effects but none of these other effects.

A disadvantage of the Solomon Four Group Experiment is its high cost

Exhibit 7.5
Diagram Illustrating a Solomon Four Group Experiment

```
        ┌──> (Experimental Group 1)  O1  X  O2
        ├──> (Control Group 1)       O3     O4
R: >────┤
        ├──> (Experimental Group 2)      X  O5
        └──> (Control Group 2)              O6
_____________________________________________

"R" indicates random assignment of subjects.
"O" indicates measurement of the dependent variable.
"X" indicates exposure of subjects to the experimental
    treatment.
```

relative to alternative experimental designs. This higher cost is due to the larger number of required groups of subjects. Four groups are necessary to test only one variable value. Given that marketing decisions almost always involve more than one independent variable and/or independent variable values, the subjects required increase accordingly by a multiple of four. Consider the Clearasil acne medication experiment, which could include numerous values/levels associated with multiple independent variables such as the product (clear, skin color, tan color, etc.), price ($1.99, $2.49, $2.99, etc.), promotion (humorous commercial, serious commercial, two versus one-sided appeal, 15-second versus 30-second, etc.), and distribution (discount stores, department stores, pharmacy, etc.).

Managerial Implication

The Solomon Four Group Experiment can serve as an ideal model when designing experiments. However, the precise information it can provide may be unnecessary and unaffordable in marketing research. Knowing the amount of effects caused by extraneous variables (also possible with Pretreatment–Posttreatment Control Group Experiments) is desirable but not necessary, since its absence does not invalidate the experiment. As discussed with the Posttreatment Only Control Group Experiment, lack of this knowledge is not the same as lack of control, since random assignment distributes extraneous variable effects equally between groups. Therefore, to accommodate best the complex nature of marketing decisions, combined with constraints of time and money, marketing research most often relies first on the Posttreatment Only Control Group Experiment and second on the Pretreatment–Posttreatment Control Group Experiment.

Also, given the complexity of marketing management, the list of independent variables related to marketing questions is practically unlimited. This fact is a serious problem for conducting a Solomon Four Group Experiment, since an additional four groups of subjects are required for each additional variable value tested. Therefore, despite the research sophistication of the Solomon Four Group Experiment, it is used relatively infrequently in marketing research. Its greatest value for marketing research is probably as an ideal to be considered when designing experiments. In conclusion, in an attempt to better accommodate the complex nature of marketing decision making, combined with constraints in time and money, marketing research more often relies on the first two types of experiments discussed in this section (i.e., the Posttreatment Only Control Group and Pretreatment–Posttreatment Control Group).

Chapter 8

Subject Assignment Category

Experiments identified in the context of subject assignments represent a choice of procedures to be used with other types of experiments. Their focus is on the randomization process, or lack thereof, that is used to assign subjects to experimental treatments. Therefore, the three experiments in this category—Completely Randomized, Randomized Block, and Latin Square—differ according to the extent to which subjects are assigned randomly to experimental groups. The assignment can be performed through a completely random process (characteristic of the Completely Randomized Experiment) or a less random process (characterized by the Randomized Block Experiment and the Latin Square Experiment). To a large extent, experiments with less randomization are generally more complex to conduct and, in general, randomization is a desirable procedure for all types of experiments.

COMPLETELY RANDOMIZED EXPERIMENT

Randomization is a fundamental aspect of experiments. However, the extent of randomization varies from complete, referred to as a completely randomized experiment, to partial, referred to as a randomized block experiment. A Completely Randomized Experiment uses a random process first, to assign subjects to groups and second, to assign groups to experimental treatments.

The objective of a Completely Randomized Experiment is to control all extraneous variables by distributing their effects equally. An underlying assumption is that random assignment distributes extraneous variable effects equally between experimental groups. A Completely Randomized Ex-

periment represents an aspect of designing experiments that is applicable to all types of experiments discussed thus far, including the factorial experiment.[1]

Random process is a procedure to select objects, in which all objects (which in experimental research are subjects and treatments) have an equal chance to be selected. With a random process there is no pattern in which individual objects are selected. A common two-step procedure to perform a random process is, first, to assign a number to each object and second, to locate any position on a random numbers table to determine the order of selecting the objects for the desired purpose.

RANDOMIZED BLOCK EXPERIMENT

For some marketing research projects, it is better to design an experiment with partial use of a random process than complete use of it. These projects are when an extraneous variable (i.e., an outside influence) is expected to influence the results of an experiment. The appropriate type of experiment in these situations is a randomized block. A Randomized Block Experiment uses a nonrandom process first, to assign subjects to groups based on a similarity factor (i.e., form blocks of subjects) and second, uses a random process to assign the groups to the experimental treatments. It requires that (1) each treatment occurs at least once within each block, and (2) at least as many groups within each block of subjects are formed as there are experimental treatments.

A distinguishing characteristic of the Randomized Block Experiment is that before experimental treatments are randomly assigned to groups of subjects, subjects are first assigned to groups based on some factor of similarity. Factors[2] may be age of subject, sex, type of store, territory, type of product, and so on. This initial assignment is always performed on a nonrandom process. The nonrandom process is used because of an expectation for a particular extraneous variable to impact responses in some consistent pattern. These responses may involve reactions to the independent variable treatment and/or the dependent variable measures. Regardless, the extraneous variable is expected to exert or account for a substantial amount of variance in the data.

With a Randomized Block Experiment, the researcher can account for at least one source of variance by blocking it out. The assumption is that through this nonrandom assignment, respective blocks are comprised of subjects whose responses to the dependent variable, even without an experimental treatment, will be more similar to each other than if subjects are randomly assigned.[3] Factors on which blocks are based may be age of subject, sex, type of store, territory, type of product, or any other variable that impacts responses in a consistent pattern and accounts for a substantial amount of variance in the data.[4]

Randomized Block Experiments involve at least three requirements. First, experimental treatments must be assigned to the individual subjects on a random basis. Second, each treatment must occur at least once within each block. Third, it is necessary to form at least as many groups within each block of subjects as there are experimental treatments, in order to meet the requirement that each treatment occurs at least once in each block. Proper procedure is to, within each block of subjects, randomly assign individual subjects to groups and then randomly assign these groups to experimental treatments.

Consider a 2×2 factorial experiment for Clearasil acne medication. The independent variables are package color (with values of "black" likely preferred by boys and "pink" likely preferred by girls) and package shape (with values of "round" intended to be carried in a purse and "flat" intended to be carried in a pant pocket). Subjects are high school sophomores. A reasonable expectation is that girls would respond more similarly in one direction to these experimental treatments and boys would respond more similarly in another direction. Therefore, subject gender is a reasonable factor to use as a blocking variable. Girls then comprise one block of subjects and boys another block of subjects, because it is reasonable to expect that girls would respond similarly to these experimental treatments, as would boys.

Three requirements were indicated above for Randomized Block Experiments: (1) The number of groups formed within each block must be at least as many as the number of experimental treatments, (2) subjects must be randomly assigned to these groups, and (3) experimental treatments must be randomly assigned to the groups while ensuring that all experimental treatments are administered at least once in each block. For the current Clearasil acne medication experiment, subjects in the block of girls are randomly assigned to four groups and then each of these four groups are randomly assigned to an experimental treatment involving either (1) a blue round package, (2) a blue flat package, (3) a pink round package, or (4) a pink flat package. Likewise, subjects in the block of boys are randomly assigned to four groups and then randomly assigned to one of the four experimental treatments. The result is eight experimental groups, as shown in Exhibit 8.1.

Objective

The assumption is that the blocking variable is correlated with the dependent variable. Therefore, the objective of a randomized block experiment is to group subjects that are more similar to each other than would result if groups are formed through a random process.

Exhibit 8.1
Prototype Table Illustrating a 2×2 Factorial Experiment with One Blocking Variable (Gender)

Table # [actual number to be provided by researcher]

Intent to Purchase According to Package Color and Package Shape with Subject Gender as a Blocking Variable

	Package Color			
	Black		Pink	
	Shape		Shape	
Block	Round	Flat	Round	Flat
Male	*	*	*	*
Female	*	*	*	*

*Each of the 8 cells represents a different experimental group of subjects, for which basic data might include the mean value of responses to an "intent to purchase" question, standard deviation value, and number of subjects.

Benefit

A benefit of a Randomized Block Experiment is increased efficiency compared to a Completely Randomized Experiment. By accounting for one type of extraneous variable and blocking its effects, a more efficient experiment is possible, since experimental error is reduced with the same sample size. This reduction is possible because some of the variation associated with the dependent variable is allocated to the blocking error. The result is less sampling error.

This less sampling error is the reason for blocking in an experiment or stratifying in a sample. In this way, a Randomized Block Experiment is analogous to stratified random sampling. Both design procedures are conducted to reduce sampling error to a level lower than possible with a completely randomized block experiment or a total random sample. In both situations it is accomplished by forming groups so that the dependent variable is more similar within groups of subjects than would occur if groups were formed randomly.

The increased research efficiency through decreased sampling error is especially apparent in the data analysis stage of the marketing research process. At that point, the differences between blocks can be accounted for in the analysis of variance. As a result, error (specifically the error mean square) for the same number of measures is smaller than it is with a completely randomized design. To obtain a similarly low level of error without a randomized block design requires a larger number of subjects.

Design Applications

A Randomized Block Experiment is an aspect of experimental design that is applicable to most all types of experiments. These are all the experiments discussed in the other categories, including the factorial experiment.[5] In almost all cases it is an important consideration in the types of quantitative analyses that are later performed on the collected data.

Variations

Typically, only one factor is utilized in a random block design. But more than one is possible. However, using more than one factor increases an experiment's complexity immensely. With each blocking factor that is *added* to an experimental design, the required number of experimental groups is *multiplied*. For example, in the above experiment, one blocking variable (gender) was used. As Exhibit 8.1 indicates, with a 2 × 2 factorial experiment, the number of groups required is eight. But only four groups would have been required if no blocking variable was used. On the other hand, assume a second blocking variable is used, such as income level of parents (high and low). With this addition of one blocking variable, the necessary number of experimental groups is multiplied by two. As shown in Exhibit 8.2, the consequence is that instead of 8 groups of subjects, 16 groups are required.

As shown in Exhibit 8.1 and Exhibit 8.2, by adding one blocking variable the number of required subjects is multiplied by two. That example, for a 2 × 2 Randomized Block Experiment, showed that increasing the number of blocking variables from one to two increased the number of groups of subjects from 8 to 16. The same doubling consequence would occur regardless of the number of independent variable treatments. However, in some limited situations, it is possible to utilize a type of experiment that does not require the large number of subjects indicated here. That type of experiment, called a Latin Square Experiment, still permits the use of two blocking variables.

Exhibit 8.2
Prototype Table Illustrating a 2 × 2 Factorial Experiment with Two Blocking Variables (Gender and Parents' Income)

Table # [actual number to be provided by researcher]

Intent to Purchase According to Package Color and Package Shape with Gender of Subjects and Income of Parents as Blocking Variables

	Package Color			
	Blue		Pink	
	Shape		Shape	
Block	Round	Flat	Round	Flat
Male High Income	*	*	*	*
Female High Income	*	*	*	*
Male Low Income	*	*	*	*
Female Low Income	*	*	*	*

*Each of the 16 cells represents a different experimental group of subjects, for which basic data might include the mean value of responses to an "intent to purchase" question, standard deviation value, and number of subjects.

Managerial Implication

A Randomized Block Experiment is an aspect that is applicable to most types of experiments,[6] and impacts quantitative analysis procedures performed later on the collected data. While substantially more subjects are required as the numbers of variables investigated increase, it is still more efficient than a Completely Randomized Experiment. This efficiency occurs because experimental error (specifically sampling error) is reduced with the same size sample, through controlling designated extraneous variables by blocking their effects.

This increased efficiency is again apparent in the data analysis, where differences between blocks can be accounted for in the analysis of variance. As a result, error (specifically the error mean square) for the same number of measures is smaller than it is with a completely randomized design. To obtain a similarly low level of error without a randomized block design requires a larger number of subjects.

Randomized Block Experiments are analogous to stratified random sampling in the context that both activities are conducted to reduce sampling error to a level lower than possible with completely randomized block experiments or total random samples. It is accomplished in both situations by forming groups so that the dependent variable is more similar within groups of subjects than would occur if groups were formed randomly.

LATIN SQUARE EXPERIMENT

A Latin Square Experiment explicitly controls two or more extraneous variables, and requires the number of values of each to equal the number of experimental treatments. Its procedure begins with random assignment of subject groups to experimental treatments that allows explicit control of extraneous variables, which is similar to other types of experiments where groups of subjects are assigned to experimental treatments through a random process. The major advantage of a Latin Square Experiment is its explicit control while requiring fewer subjects than alternative experiments. A 2 × 2 × 2 factorial experiment requires eight subject groups to investigate three different variables with two values each. But the same variables and values can be investigated with a 2 × 2 Latin Square Experiment with four groups.

Before discussing differences, it is valuable to note that the Latin Square Experiment has similarities with other types of experiments. For example, like a Randomized Block Experiment, a Latin Square Experiment is appropriate when it is desirable to explicitly control extraneous variables. These times are, of course, when specific extraneous variables are expected to influence the data in a consistent pattern. Also, like a factorial experiment utilizing a completely randomized design, a Latin Square Experiment permits simultaneous investigation of more than one variable. For example, while a 2 × 2 × 2 factorial experiment refers to investigating three different variables, each with two levels, the same variables and values can be investigated with a 2 × 2 Latin Square Experiment.

The major advantage of a Latin Square Experiment is its explicit control while requiring fewer subjects than alternative experiments such as a Randomized Block Experiment or a factorial experiment in the design of a Completely Randomized Experiment. A 2 × 2 × 2 factorial experiment requires eight subject groups to investigate three different variables with two values each. But the same variables and values can be investigated with a 2 × 2 Latin Square Experiment with four groups.

Exhibit 8.3
Randomized Block Experiment to Investigate Three Variables

Table # [actual number to be provided by researcher]

Intent to Purchase According to Package Color with Gender of Subjects and Income of Parents as Blocking Variables

	Package Color	
Block	Blue	Pink
Male High Income	*	*
Female High Income	*	*
Male Low Income	*	*
Female Low Income	*	*

*Each of the 8 cells represents a different experimental group of subjects, for which basic data might include the mean value of responses to an "intent to purchase" question, standard deviation value, and number of subjects.

To illustrate the need for fewer subjects, compare the three experimental designs, each investigating three variables, presented in Exhibit 8.3, Exhibit 8.4, and Exhibit 8.5. The Randomized Block Experiment and the factorial experiment require eight groups of subjects, while the Latin Square Experiment requires only four. However, while more than one extraneous variable can be investigated with a Randomized Block Experiment, only the overall block effect can be determined. For example, any effects indicated in Exhibit 8.3 could not be separated according to gender of subjects and income of parents.

Exhibit 8.4
Factorial Experiment to Investigate Three Variables

Table # [actual number to be provided by researcher]

Intent to Purchase According to Package Color, Income of Parents, and Gender of Subjects

	Package Color			
	Blue		Pink	
	Parents' Income		Parents' Income	
Sex	High	Low	High	Low
Male	*	*	*	*
Female	*	*	*	*

*Each of the 8 cells represents a different experimental group of subjects, for which basic data might include the mean value of responses to an "intent to purchase" question, standard deviation value, and number of subjects.

Each of these three designs represents an investigation of three variables. Both the Randomized Block Experiment and the factorial experiment require eight groups of subjects, and the Latin Square Experiment requires only four. Furthermore, while more than one extraneous variable can be investigated with a Randomized Block Experiment, when there are more than one such variables, the researcher can only determine the overall block effect. For example, any different effects indicated through an experiment as illustrated in Exhibit 8.3 could be due to either gender of subjects and/or income of parents. Separate effects cannot be determined in this type of experiment.

Disadvantages

The advantage of fewer required subjects is accompanied by severe disadvantages for use of the Latin Square Experiment design in marketing research. First, interaction effects between independent variables and in-

Exhibit 8.5
Latin Square Experiment to Investigate Three Variables

Table # [actual number to be provided by researcher]

Intent to Purchase According to Income of Parents and Gender of Subjects

	Parents' Income	
Gender	High	Low
Male	* "A"	* "B"
Female	* "B"	* "A"

*Each of the 4 cells represents a different experimental group of subjects, for which basic data might include the mean value of responses to an "intent to purchase" question, standard deviation value, and number of subjects.

"A" and "B" represent Experimental Treatments for Package Color (with two values or levels):

"A" represents Blue Package
"B" represents Pink Package

dependent variable values cannot be investigated. Therefore, a Latin Square Experiment is limited to those situations in which it is reasonable to assume that no interaction effects occur.

Second, independent variable values (i.e., experimental treatments) must be equal in number to the number of values for each extraneous variable explicitly controlled. Hence, the "square" name of this type of experiment is reflected in the corresponding "square" table that represents its design. These equal numbers are a requirement. However, in the real world of marketing, the number of independent variable values varies, as do the values of categories of the extraneous variables. While the above experimental design reflects a 2 × 2 Latin Square Experiment with two values of an independent variable, it could just as well be a 4 × 4 Latin Square Experiment with four independent variable values, or a 3 × 3 with three values. The corresponding design diagram would, instead of being a 2 × 2 square or matrix, be a 4 × 4 or 3 × 3. The only requirement is an equality that translates into a square diagrammed design.

Managerial Implication

The advantage of fewer required subjects is accompanied by severe disadvantages for use of the Latin Square Experiment design in marketing research. First, interaction effects between independent variables and independent variable values cannot be investigated. Therefore, a Latin Square Experiment is limited to those situations in which it is reasonable to assume that no interaction effects occur, which in the reality of a marketing decision maker is very unlikely.

Second, independent variable values (i.e., experimental treatments) must be equal in number to the number of values for each extraneous variable explicitly controlled. Hence, the "square" name of this type of experiment is reflected in the corresponding "square" table that represents its design. However, the real world of marketing decision makers often involves situations where the number of independent variable values varies differently than the values of extraneous variables. For example, questions about gender might involve two values while questions about package design involve two, three, four, or more values.

Furthermore, there are serious trade-offs when using a Latin Square Experiment for a marketing research project. While a positive aspect is fewer required subjects, a negative aspect is the numerous limitations determining when it can be used. First, it is practical only with one independent variable. Second, there must always be an equal number of independent variable values and extraneous variable values. Third, there is no test of interaction effects. In marketing, the reality is that there is most always an interest in more than one independent variable and there are most always interaction effects between variables. Therefore, despite the positive appeal of less subjects, which means less time and money, the negative side is a serious limitation as to when this type of experiment can be used in marketing research.

NOTES

1. There is apparently a discrepancy on this issue among marketing research textbooks. Some state that a completely randomized design is applicable to experiments with more than one independent variable, such as the factorial experiment (e.g., see Gilbert A. Churchill, Jr., *Basic Marketing Research* [Chicago: Dryden Press, 1988], p. 613). But others state this design is applicable only to experiments with one independent variable (e.g., see William G. Zikmund, *Exploring Marketing Research*, 3d ed. [Chicago: Dryden Press, 1989], p. 323, and Thomas C. Kinnear and James R. Taylor, *Marketing Research*, 3d ed. [New York: McGraw-Hill,

1987], p. 350). Theoretically and realistically, there is no justification for limiting application of a completely randomized experiment to experiments with only one independent variable.

2. For a thorough consideration of factors see Seymour Banks, "Experimental Design in Control," in *Handbook of Marketing Research*, Robert Ferber, ed. (New York: McGraw-Hill, 1974), Section 2, p. 475.

3. See Allen L. Edwards, *Experimental Design in Psychological Research*, 3d ed. (New York: Holt, 1968), p. 155.

4. For a thorough consideration of factors see Banks, "Experimental Design in Control," Section 2, p. 475.

5. See note 1.

6. See note 1.

Chapter 9

Quasi-Experiments

The word "experiment" is often used in a loose and technically incorrect manner. In these instances it refers to "trying something and seeing what happens." Such usage of the word only crudely and remotely resembles the cause-and-effect aspect of an actual experiment. The difference is that with an actual experiment, there are systematic procedures to assure proper control. In turn, this control assures that measures are properly assessed, against which baseline measures (i.e., norms) are compared. The result is that an actual experiment allows conclusions to be made with reasonable confidence about "what actually happened when something was tried."

While an actual experiment may be the ideal, realities of marketing research often pose constraints. These constraints necessitate research efforts that are somewhat less than actual experiments. At these times it is necessary to approximate the ideal of an experiment, while omitting aspects judged too costly in terms of time and money. This type of research project is less complex and less sophisticated than an actual experiment, and is referred to as a "quasi-experiment."[1]

A quasi-experiment is a research project that approximates an experiment, but lacks the control that permits comparison of effects possible with an experiment. Quasi-experiments are not necessarily bad, as long as researchers and users of the results keep aware of the major limitation. This limitation is that conclusions about cause-and-effect relationships based on the results from quasi-experiments cannot be made with the same level of certainty as when procedures and controls of an experiment are employed.

A quasi-experiment is similar to an experiment in terms of the hypothesis, independent variable, dependent variable, and subjects. It is typically dissimilar in terms of the control procedures involving manipulation of

experimental treatments and consideration of extraneous variables. Conclusions about causal relationships between independent and dependent variables are therefore more likely to be less accurate.

Actual experiments involve a minimum of two groups of subjects. One group, the experimental group, receives an initial measure of the dependent variable, followed by an experimental treatment, followed by a second measure of the dependent variable. The other group, the control group, receives an initial measure, does *not* receive an experimental treatment but does allow the same period of time to pass, followed by a second measure. These procedures permit comparison of measures for the experimental group against measures for the control group which provide baseline measures or norms. The difference between the two groups can then be inferred to be the effect on the dependent variable caused by the independent variable. Quasi-experiments commonly omit this control group and/or one or more of the measures. Either omission technically disqualifies it as an experiment, which means that conclusions must be qualified accordingly.

TYPES OF QUASI-EXPERIMENTS

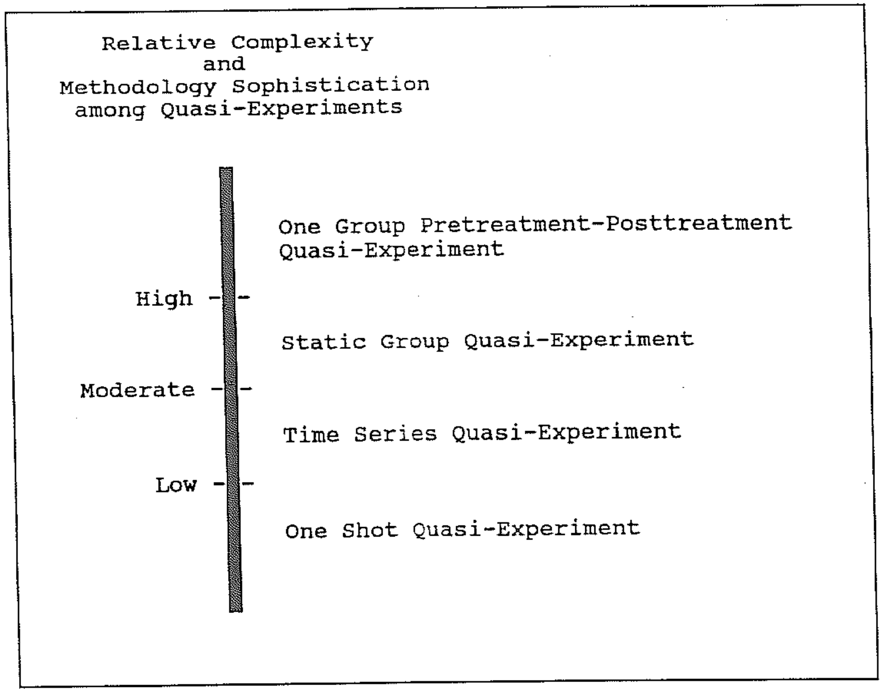
Although results from quasi-experiments must be qualified, due to procedures that lack control and comparison measures, quasi-experiments are not without value. They can provide information, albeit limited. It fact, a quasi-experiment may at times be the most appropriate research alternative given a company's defined research problem, its budget limitations, and its time constraints. That is, most appropriate as long as the results are qualified that any related conclusions cannot be made with the same level of certainty that is possible with an actual experiment.

Four types of quasi-experiments are frequently used in marketing research: One Group Pretreatment–Posttreatment, Static Group, One Shot, and Time Series.[2] The complexity and sophistication of the methodology of these four quasi-experiments is not the same, and as indicated in Exhibit 9.1, can be viewed as ranging along a continuum. Furthermore, each of these quasi-experiments can in some sense be viewed as a modified or "stripped down" version of the experiments discussed earlier.

One Group Pretreatment–Posttreatment Quasi-Experiment

Of the quasi-experiment types, the One Group Pretreatment–Posttreatment Quasi-Experiment most closely parallels the standard experiment identified earlier as the Pretreatment–Posttreatment Control Group Experiment. The One Group Pretreatment–Posttreatment Quasi-Experiment approximates an experiment, but lacks the control that permits comparison of effects possible with an experiment. Specifically, it is a quasi-experiment in which a measure of the dependent variable is assessed before and after

Exhibit 9.1
Relative Complexity and Methodology Sophistication of Four Quasi-Experiments



exposure to the experimental treatment. Exhibit 9.2 illustrates how the One Group Pretreatment–Posttreatment Quasi-Experiment differs from the standard Pretreatment–Posttreatment Control Group Experiment. The key difference is that no control group is used in this type of quasi-experiment. This design aspect limits internal validity due to corresponding lack of control of any extraneous variable effects due to history, mortality, testing, and so on.

Decisions pertaining to Clearasil acne medication offer a marketing research example of the One Group Pretreatment–Posttreatment Quasi-Experiment. The independent variable is package size. The hypothesis is that a large (8-ounce) size would cause more sales (dependent variable) than the current (2-ounce) size. Denver is selected as the place in which to test this hypothesis. The current sales level of the product in Denver is assessed, distribution of the new 8-ounce package size is made in the city for three months, after which a second measure of the product sales is assessed. Subtracting the second measure from the first measure provides an indication of the experimental treatment effects (i.e., the impact of the 8-ounce package size on sales). However, the certainty of the conclusion about these effects is limited, since there was no control group against which to compare the measures. As a result, any apparent difference could have been caused by extraneous variable effects rather than, or in addition to, treatment effects.

Exhibit 9.2
Diagram of a One Group Pretreatment–Posttreatment Quasi-Experiment

```
Diagram illustrating a basic Pretreatment-Posttreatment
Control Group Experiment:

          ┌──> (Experimental Group)    O1   X   O2
R:  >─────┤
          └──> (Control Group)         O3       O4

_____________________________________________

Diagram illustrating a basic One Group Pretreatment-
Posttreatment Quasi-Experiment:

               (Experimental Group)    O1   X   O2

_____________________________________________

Diagram illustrating a variation of the One Group
Pretreatment-Posttreatment Quasi-Experiment in which
more than one experimental group is used:

          ┌──> (Experimental Group 1)  O1   X   O2
R:  >─────┼──> (Experimental Group 2)  O3   X   O4
          └──> (Experimental Group 3)  O5   X   O6

_____________________________________________

"R" indicates random assignment of subjects.
"O" indicates measurement of the dependent variable.
"X" indicates exposure of subjects to the experimental
    treatment.
```

Multiple Experimental Groups. In its most basic design there is only one treatment group in the One Group Pretreatment–Posttreatment Quasi-Experiment. But more than one treatment group can be used. In such a variation, as Exhibit 9.2 shows, there is random assignment of subjects but there still is no control group. By subtracting the respective before and after measures (i.e., $O_1 - O_2$ for treatment one, $O_3 - O_4$ for treatment two, and $O_5 - O_6$ for treatment three), the resulting amounts could be compared to suggest relative effectiveness of the three different treatments (i.e., the different effects of the three different package sizes on sales of the product).

Instead of one new package size in the above Clearasil acne medication example, three new package sizes (4 ounces, 8 ounces, and 16 ounces) might be tested, respectively, in Charlotte, North Carolina; Kansas City, Missouri; and Sacramento, California. Calculating the respective sales differences would suggest which package size has the greatest and least effect on the dependent variable. Even with apparent differences, there is of course still concern about internal validity due to the lack of a control group to serve as a norm or baseline measure with which these observed differences can be compared.

Managerial Implication. Despite the limitation, this type of quasi-experiment is used frequently in marketing research. While its weakness is internal validity, its strength is its ability to indicate treatment effects, through premeasures and postmeasures, at relatively low cost.

Static Group Quasi-Experiment

A Static Group Quasi-Experiment is a quasi-experiment in which a measure of the dependent variable is assessed after exposure to the experimental treatment and a control group is included. Treatment effects are calculated by subtracting the measure of the experimental treatment group from the measure of the control group (i.e., $O_1 - O_2$ as shown in Exhibit 9.3). Technically, the designated control group in Exhibit 9.3 is simply a second group. However, in marketing research it is reasonable to consider this second group as a control group, since it is often defined in terms of subjects receiving the current level of marketing effort. It can therefore be used for baseline measures, to make comparisons with the experimental group, which is comprised of subjects receiving the "new" level of marketing effort.

As Exhibit 9.3 illustrates, the design diagram of the Static Group Quasi-Experiment is almost identical to the type of experiment identified earlier as the Posttreatment Only Control Group Experiment. The key difference is that with this quasi-experiment a random process is not used to assign subjects to the groups. This design aspect limits internal validity due to corresponding lack of control regarding extraneous variable effects, due, for example, to selection and mortality.

Because of possible selection effects, differences in observed measures on the dependent variable could be due to the fact that the groups differed even before exposure to the experimental treatment. Because of possible mortality effects, any observed differences could be due to more subjects withdrawing from the experimental group than from the control group. A reason for this difference of withdrawal rates might be some unpleasantness

Exhibit 9.3
Diagram of a Static Group Quasi-Experiment

```
Diagram illustrating a basic Posttest Only Control Group
Experiment:

         ┌──> (Experimental Group)  X  O1
R:  >────┤
         └──> (Control Group)          O2

______________________________________________

Diagram illustrating a basic Static Group Quasi-
Experiment:

          (Experimental Group)      X  O1
          (Control Group)              O2

______________________________________________

Diagram illustrating a variation of the Static Group
Quasi-Experiment:

          (Experimental Group 1)      X  O1
          (Experimental Group 2)      X  O2
          (Experimental Group 3)      X  O3
          (Control Group)                O4

______________________________________________

"R" indicates random assignment of subjects.
"O" indicates measurement of the dependent variable.
"X" indicates exposure of subjects to the experimental
    treatment.
```

associated with the treatment. Omitting pretreatment measures makes it impossible to determine the extent of such extraneous variable effects.

A Static Group Quasi-Experiment is particularly applicable in situations where a comparison is being made between the status quo and a proposed change. Consider again the Clearasil acne medication experiment to assist decision making about the impact on sales likely to be caused by one new, larger-sized package. By assuming that customers for this product are similar in two different cities (e.g., Denver and Kansas City), a Static Group Quasi-Experiment could be conducted at minimal cost. After, let's say, three months of marketing Clearasil in the new package size in Denver, the sales data from Denver could be compared with sales data from Kansas City, where the product was marketed continually in its traditional/historical package size. The limitation is, of course, that typical uncontrollable environmental variables (economy, competition, etc.) may be different in the two cities and their influence on the sales data is unknown.

Multiple Experimental Groups. In its most basic design, there is only one treatment group in the Static Group Quasi-Experiment. But more than one treatment group can be used. In such a variation, as Exhibit 9.3 illustrates, there are multiple treatment groups with still no pretreatment measures. By subtracting the respective after-treatment measures (i.e., $O_1 - O_4$ for treatment one, $O_2 - O_4$ for treatment two, and $O_3 - O_4$ for treatment three), the resulting amounts could be compared to suggest relative effectiveness of the three different treatments.

Instead of one new package size in the above example, three new package sizes (8 ounces, 12 ounces, and 16 ounces) might be tested in different cities. Calculating the respective sales differences (i.e., comparing them against the baseline or norm level represented by the control group measure O_4) would suggest which package size has the greatest and least effect on the dependent variable. Of course, even with such a difference, there are still concerns about internal validity, due to the lack of a pretreatment measure group that would serve to assure the groups were all equal initially, before any experimental treatments were administered.

Managerial Implication. Despite the limitation, this type of quasi-experiment is frequently used in marketing research. While its weakness is internal validity, its strength is its ability to indicate treatment effects (via a control group) at relatively low cost. Furthermore, in some instances it may simply not be possible to administer a pretreatment measure. For example, consider trying to conduct a marketing research project for a new product yet to be introduced. Consumers in this situation have had no experience or use of the product, which could help provide a meaningful basis on which to assess a pretreatment measure.

One Shot Quasi-Experiment

A One Shot Quasi-Experiment is a quasi-experiment comprised entirely of only one experimental group on which only a posttreatment measure is assessed. As Exhibit 9.4 illustrates, there are no pretreatment measures and no control group measures. An example of a One Shot Quasi-Experiment is provided in relation to the previous Clearasil acne medication research. The research procedure is to market the product in a larger, 16-ounce package, for three months, and then measure sales.

Marketing researchers use this methodology more frequently than might be generally expected. However, its problem, technically, is that conclusions about the posttreatment measure are largely speculation. The reason is that effects of extraneous variables are completely uncontrolled and no control group measure exists.

Disadvantage. There is a problem in this situation with making conclu-

Exhibit 9.4
Diagram of a One Shot Quasi-Experiment

```
Diagram illustrating a basic One Shot Quasi-Experiment:

          (Experimental Group)        X  O1

_____________________________________________

Diagram illustrating a variation of the One Shot
Quasi-Experiment:

          (Experimental Group 1)        X  O1
          (Experimental Group 2)        X  O2
          (Experimental Group 3)        X  O3

_____________________________________________

"O" indicates measurement of the dependent variable.
"X" indicates exposure of subjects to the experimental
     treatment.
```

sions related to treatment effects such as the above suggestion to measure sales in regard to testing impact of a new package size. The problem is that it is almost complete speculation to make conclusions about whether the posttreatment measure of sales indicated the experimental treatment was successful or not. This problem is due to possible effects caused by extraneous variables and the lack of baseline or norm measures for comparisons.

Of course, somewhat more information to help address this problem could be provided through a variation that uses multiple experimental groups, as shown in Exhibit 9.4. Respective calculations are then a simple ranking of the different sales measures (O_1, O_2, and O_3). While there still is no baseline or norm measure for comparison purposes, at least the comparison of sales figures suggests which treatment was most effective.

Managerial Implication. Of the three different quasi-experiments discussed, the One Shot Quasi-Experiment is the weakest in terms of internal validity. It possesses all the concerns expressed in regard to *both* the One Group Pretreatment–Posttreatment Quasi-Experiment and the Static Group Quasi-Experiment (e.g., extraneous variable effects due to history, mortality, testing, selection, etc.). At the same time, its strength is its low cost in terms of time and money to conduct the research. Also, in situations where a pretreatment measure is not practical (e.g., a new product or a taste test) the One Shot Quasi-Experiment may be an acceptable choice among the quasi-experiment alternatives.

Time Series Quasi-Experiment

Neither the experiments nor quasi-experiments discussed up to this point focus on trends from one time period to another. Yet company executives often have questions and hypotheses about changes that occur over time. For example, they may wish to track marketplace attitudes about a product during a year-long time period. To do this tracking, marketing research involves periodic measures (such as once a month during a specified year). By doing this rather long-term tracking, longer trends in the marketplace can be identified as well as shorter-term changes. The methodology to address this situation, to track attitudes or sales in order to document the long-term effects of experimental treatments is formally referred to as a Time Series Quasi-Experiment. A Time Series Quasi-Experiment is a quasi-experiment with repeated measures of the dependent variable over an extended period of time.

As Exhibit 9.5 illustrates, a Time Series Quasi-Experiment lacks an essential aspect of an experiment: It does not control for effects of extraneous variables. While its repeated measures do permit some control of maturation effects, they do not permit control of effects due to history, mortality, interactive testing, and selection. A way, however, to increase control of the extraneous variable effects is the variation with control group, also shown in Exhibit 9.5. For illustration of a Time Series Quasi-Experiment, a related marketing research project is described in Exhibit 9.6.

By definition, Time Series Quasi-Experiments are conducted over an extended period of time. They are therefore a type of longitudinal study. Generically, a longitudinal study is a study with repeated measures, over time, of the same sample of elements or subjects. It might be noted that Time Series Quasi-Experiments are more commonly referred to as "time series studies." For practical purposes, time series study is synonymous with the Time Series Quasi-Experiment and, in reality, the two terms can be used interchangeably.

Panels as Subjects. Time series studies conducted in marketing research often utilize individuals who comprise a marketing research panel. A panel is a group of subjects who are involved in a marketing research study over an extended period of time. This period of time can range from weeks to years. There are two major categories or types of panels: the traditional and the omnibus. The traditional panel is a panel in which its members are measured repeatedly on the same variables. For this panel, the same questions are repeated to the same people over a length of time.

A variation of the traditional panel is the omnibus panel. The omnibus panel is a panel in which its members are measured repeatedly but *not*

Exhibit 9.5
Diagram of a Time Series Quasi-Experiment

Diagram illustrating a basic Time Series Quasi-Experiment:

(Experimental group)	O_1	O_2	O_3	X	O_4	O_5	O_6

Diagram illustrating a variation of a Time Series Quasi-Experiment with Control Group:

(Experimental group)	O_1	O_2	O_3	X	O_4	O_5	O_6
(Control group)	O_7	O_8	O_9		O_{10}	O_{11}	O_{12}

Diagram illustrating a variation of a Time Series Quasi-Experiment with Multiple Experimental Groups:

(Experimental group 1)	O_1	O_2	O_3	X	O_4	O_5	O_6
(Experimental group 2)	O_7	O_8	O_9	X	O_{10}	O_{11}	O_{12}
(Experimental group 3)	O_{13}	O_{14}	O_{15}	X	O_{16}	O_{17}	O_{18}

"O" indicates measurement of the dependent variable.
"X" indicates exposure of subjects to the experimental treatment.

necessarily on the same variables. For this panel, the questions may differ from measure to measure but they are typically within the same general product category, such as alcoholic beverages, automobiles, baby products, and so on.

Panels, regardless of type, pose questions about internal validity related to most every extraneous variable effect discussed up to this point. Especially problematic are selection effects. People who serve on panels are largely self-selected and likely to have different characteristics which allow them to participate in a study for an extended period of time. Likewise, the long-term aspect of participating in panels makes effects due to extraneous variables such as mortality, maturation, history, and testing nearly impossible to avoid. Despite these problems, panels are commonly used to

Exhibit 9.6
Marketing Research Project Involving a Time Series Quasi-Experiment: A Firsthand Experience Example

A marketing research project conducted by the author of this book provides an example of a standard, thorough, Time Series Quasi-Experiment. The organization (American Savings and Loan Association in Albuquerque) wished to measure awareness of and attitude to their company in the marketplace over a year-long time period. During that year, several "experimental treatments" were administered involving print and broadcast promotions for a cooperative advertising effort in which opening certain savings accounts would allow substantial discounts for purchasing certain major products.

During the year that measures of awareness and attitude were assessed, a promotion for a washer-dryer set was announced in May and a promotion for car tires was announced in September. The design diagram for this example was then as follows:

$O_1\ O_2\ O_3\ O_4\ O_5\ X\ \ O_6\ O_7\ O_8\ O_9\ X\ \ O_{10}\ O_{11}\ O_{12}$

This quasi-experimental design provided company executives an indication of the marketplace awareness and attitude before any special promotional efforts/experimental treatments, as indicated by the January through May measures (O_1, O_2, O_3, O_4, and O_5). With the washer-dryer promotion in May, the June through September measures (O_6, O_7, O_8, and O_9) provided an (approximate) indication of this promotion's effectiveness in influencing attitude and awareness. Likewise, with the car tires promotion in September, the October through December measures (O_{10}, O_{11}, and O_{12}) provided an indication (albeit approximate) of that promotion's effectiveness.

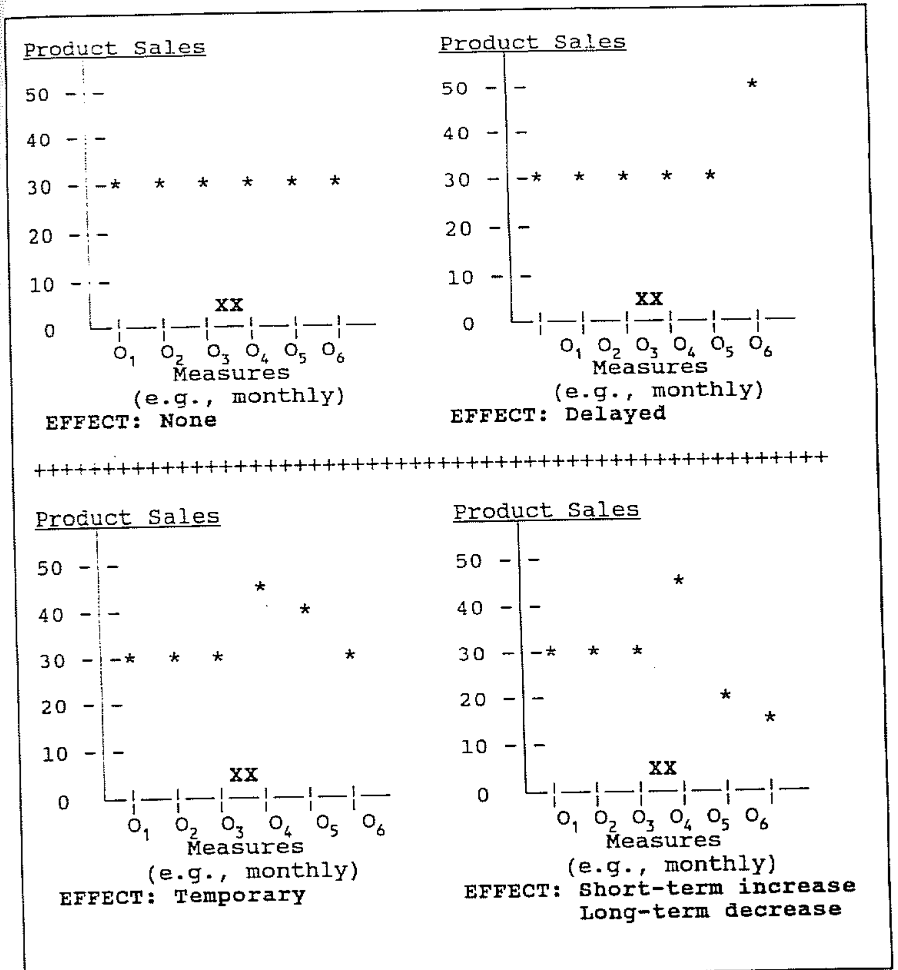
In regard to the later measures, any conclusions by company executives had to be tempered by concerns about internal validity caused by testing effects. There are also cumulative treatment effects or carryover effects, in that the two separate treatments may not have produced two separate effects. For example, the fact the second promotion (for car tires) was preceded by the first (for a washer-dryer set), could cause a greater impact on the subjects than if the second promotion was performed alone without the first.

provide information on which marketing executives base important company decisions.

These panels are comprised of subjects representing both distribution channel members and final consumers. Products for which panels exist range from business-to-business products to political campaigns to consumer products. Panels are as small as a few subjects or as large as thousands of subjects. Two of the largest are the National Purchase Diary Panel[3] comprised of 13,000 families, and the Nielsen Retail Store Audit[4] panel comprised of 10,000 stores.

Managerial Implication. A Time Series Quasi-Experiment has problems with internal validity, but it has a powerful advantage in terms of contin-

Exhibit 9.7
Four Selected Results of Time Series Quasi-Experiments



uous information over a long period of time at relatively low cost. A major value of a time series study, as illustrated in Exhibit 9.7, is its ability to provide information about long-term effects, as well as short-term effects, on a dependent variable caused by an independent variable.

For example, the measures in Exhibit 9.7 could span a period of several months of time as may be of interest to a certain company or decision maker. During this time period, a time series study may reveal quite different long-term and short-term effects of an experimental treatment. Consider an experimental treatment such as a new package size or new price strategy. If company decisions are made immediately after an experimental treatment, such as illustrated by the O_4 measure in Exhibit 9.7, these decisions would be substantially different than if the company waited a few months after the treatment and based its decisions either on the O_6 measure or the trend indicated by the O_4, O_5, and O_6 measures together.

NOTES

1. For a more generic and thorough discussion of quasi-experiments see Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Design for Research* (Chicago: Rand McNally, 1966), pp. 34–64; Seymour Banks, *Experimentation in Marketing* (New York: McGraw-Hill, 1965), pp. 37–45; James Caporaso, "Quasi-Experimental Approaches to Social Science: Perspectives and Problems," in James A. Caporaso and Leslie L. Roos, Jr., eds., *Quasi-Experimental Approaches: Testing Theory and Evaluating Policy* (Evanston, Ill.: Northwestern University Press, 1973), pp. 3–38.
2. For a more generic and thorough discussion of types of quasi-experiments, including additional types, see Campbell and Stanley, *Experimental and Quasi-Experimental Design for Research*, pp. 34–64; Seymour Banks, "Experimental Design in Control," in *Handbook of Marketing Research*, Robert Ferber, ed. (New York: McGraw-Hill, 1974), pp. 37–45; James Caporaso, "Quasi-Experimental Approaches to Social Science: Perspectives and Problems," in Caporaso and Roos, eds., *Quasi-Experimental Approaches: Testing Theory and Evaluating Policy*, pp. 11–31.
3. See publication by the National Purchase Diary Panel, *We Make the Market Perfectly Clear* (New York: National Purchase Diary Panel, Inc., undated).
4. See publication by A. C. Nielsen Company, *Management with the Nielsen Retail Index System* (Northbrook, Ill.: A. C. Nielsen Company, 1980).

Part III

Experiments in Marketing Research in Action

The third part of this book describes actual experiments in marketing research, rather than only the different dimensions and types of experiments available to a marketing researcher. Chapter 10 identifies the stages and activities required to properly conduct experiments in marketing research. Since experiments are conducted in the context of marketing research, the marketing research process is discussed, including its first two stages: problem definition and research design.

Actual experiments conducted by the author of this book to assist respective marketing decision making are described in Chapters 11 and 12. Description of the experiments here are accompanied with discussion about various procedural and analytical considerations. A major difference between these two chapters is the level of quantitative analysis conducted and described. In Chapter 11, data collected through an experiment are analyzed first through calculating and comparing percentages. The data are then analyzed through analysis of variance, which is a more sophisticated quantitative technique that is used frequently to analyze data from an experiment. In contrast, Chapter 12 includes greater rationales and explanations for various procedures and demonstrates a multitude of quantitative tests and techniques which may be performed to analyze an experiment's data.

Chapter 10

Experiments and the Marketing Research Process

Experiments offer marketing researchers the most powerful methodology for making conclusions about causality, which is the causal relationship between two or more variables. That is, if the many dimensions of experiments are considered carefully, the type of experiment design is selected properly, and the experiment is then conducted properly. Furthermore, all of these actions and decisions must be carried out within the context of the overall marketing research process, which must begin with a strong problem definition and research design.

CONDUCTING AN EXPERIMENT

Four stages of activities can be identified to depict the conducting of an experiment. As Exhibit 10.1 identifies, these stages are to: (1) formulate hypotheses to be tested in response to the problem definition stage of the overall marketing research project, (2) design the experiment, (3) perform the experiment, which includes collecting the data, tabulating, analyzing, and interpreting the data as appropriate, and (4) present a report that describes the experiment, its findings, and, often, its implications for assisting the decision for which it was conducted.

Each of the stages necessary to conduct an experiment involves a myriad of decisions to be made and activities to be performed. As this process begins, it is important to keep in mind that it only begins after the problem definition stage of the marketing research process is completed. In other words, the hypotheses are formulated in response to the problem(s) defined in the initial stage of the overall marketing research process, followed by

Exhibit 10.1
Stages Necessary to Properly Conduct an Experiment

```
(1) FORMULATE HYPOTHESES

(2) DESIGN THE EXPERIMENT

(3) PERFORM THE EXPERIMENT

(4) PRESENT A REPORT (OF THE EXPERIMENT)
```

designing and performing an experiment, and then presenting a report of the experiment.

THE MARKETING RESEARCH PROCESS

Mention has been made numerous times in this book about the problem definition stage and the research design stage of marketing research. In order to accurately portray the role of experiments as a methodology within marketing research, it is worthwhile to note their position relative to the problem definition and research design stages of the marketing research process—a series of sequential, interdependent stages required to conduct a marketing research project. Also, marketing research can be defined as systematic inquiry to provide marketing information to assist decision making.

While the activities within the marketing research process are extensive, they are commonly categorized into five stages: problem definition, research design, data collection, data analysis, and report presentation. These stages encompass only a relative few of the major activities as outlined in Exhibit 10.2. As noted in this exhibit, selection of an experiment as the appropriate methodology for a particular marketing research project is most prominent in regard to consideration of the research methodology within the research stage. Therefore, the stages of problem definition and research design are particularly relevant to conducting experiments in marketing research.

Problem Definition

Problem definition is the clear statement of the objective for conducting a marketing research project. It is the first stage of the marketing research

Exhibit 10.2
Outline of the Marketing Research Process

```
1.  PROBLEM DEFINITION  (OR DEFINE THE PROBLEM)

        Research Objective(s) and Research Questions

2.  RESEARCH DESIGN  (OR DESIGN THE RESEARCH)

        Type of Data
            Secondary and/or Primary
            Qualitative and/or Quantitative
            Syndicated, etc.
        Sampling
            Population
            Type of Sampling Method/Type of Sample
            Sampling Frame
            Sample Size
            Operational Procedures
        Data Collection
            Methodology (to access and record data)
                ✓Experiment✓, Survey, Observation
                Tools, Forms, Questionnaires
            Procedures, Personnel, Location, etc.
        Data Analysis
            Preparation, Investigation, Interpretation,
            (with or without Recommendations)
        Report Presentation
            Verbal and Written

3.  DATA COLLECTION  (OR COLLECT THE DATA)

        Access Data and Record Data

4.  DATA ANALYSIS  (OR ANALYZE THE DATA)

        Prepare, Investigate, and Interpret Data

5.  REPORT PRESENTATION  (OR PRESENT THE REPORT)

        Verbal Report and Written Report
```

process and is followed by the research design activities. Proper problem definition, in which the marketing research objectives and questions are explicitly stated, helps focus activities and avoid later misunderstandings between researchers and those who use the research to assist their decision making. Consider the vignette in Exhibit 10.3 about the National Basketball Association's decision to open an office in Hong Kong. Proper problem definition would state that:

The research objective is to investigate the potential for American basketball in China.

Related research questions expected to be answered could include:

Exhibit 10.3
American Basketball in Hong Kong

NBA (National Basketball Association) management explores new markets continually...and globally. Like most organizations, the marketing research which the NBA uses to assist its respective decisions begins with problem definition to clarify research objectives and questions to be answered.

Consider the NBA's decision to locate an office in Hong Kong. Once the problem was defined as not knowing the potential of American basketball in that area of the world, the necessary research was designed. However, marketing research in the Far East requires different procedures than marketing research in the United States. For example, telephone interviews are rare and all communication must be in the respondent's native language, unless the target market is very exclusive with highly educated or successful international business persons. Secondary data in China are nonexistent relative to countries such as the United States. Psychological and sociological dynamics that motivate respondents to give "correct" or socially desirable answers, rather than true answers, complicate the research. The consequence is often misleading, overly optimistic data. Therefore, the research design for conducting the NBA's marketing research had to counter these difficulties.

Why Hong Kong? Because marketing research revealed it is ideally located: central between Japan and Singapore, near the Philippines and Taiwan, and on the doorstep of China. Why China? Because basketball is already popular there. More than 130 million people are active basketball players, which is more than 1/2 the entire American population. Another dimension, China has over 1.2 billion people (compared to 260 million in the United States) and over 40% of the Asian population is under 35 years of age. The decision for the NBA to locate an office in Hong Kong became rather clear in light of this information provided through marketing research via proper problem definition and proper research design.

This example is based on personal experiences of the author while living in Hong Kong and travelling in China, as well as an article in the Asian Wall Street Journal: Luke Cyphers (1992), "U.S. Basketball Targets Asia as Big Market," Asian Wall Street Journal, vol. XVI, no. 153, pp. 1 & 5.

What is the population of China?

What is the age distribution of people in China?

How aware of basketball are people in China?

How many people in China like basketball?

How many people in China play basketball, and how often?

As the NBA example illustrates, the term "problem definition" is not limited to negative situations. Since problem pertains to the difficulty associ-

Exhibit 10.4
Seven Factors That Contribute to Improper Problem Definition

(1) Difficult Task

(2) Activity is Ambiguous/Unstructured

(3) Hasty Tendency to Proceed with Project

(4) Researcher Lacks Experience

(5) Inadequate User-Researcher Interaction

(6) Goals of User are Imprecise

(7) Goals of User are Unrealistic

ated with making the best decision,[1] it applies also for positive situations that pose opportunities.

Distinction of Problem Definition. Its far-reaching influence makes problem definition distinct within the marketing research process. While proper problem definition does not ensure a successful marketing research project, improper problem definition destines a project to fail. This impact is due to establishing a goal early in a research project, which then serves as a form of mission statement that guides subsequent activity by prescribing the general task to be accomplished.

Problem definition is the most important research stage for contribution from users of the marketing research. The reason is that it requires them to delineate information that will assist best their decision making. A weakness in their contribution here negatively impacts the success of the project by decreasing likelihood of receiving information which will truly assist. Therefore, eventual users of the research should know what information they need and want, communicate this information, and understand the feasibility of obtaining it.

Deterrents to Proper Problem Definition. Exhibit 10.4 identifies seven factors that contribute to improper problem definition. Imprecise and unrealistic goals of marketing research users are prominent among these factors. A common scenario is a person commissioning a marketing research project with only vague notions about its procedures, limitations, and ability to assist decision making. These scenarios end with research projects that fail to assist decision making, but which do not publicly recognize imprecise or unrealistic goals as the cause.

Researchers and those who commission the research usually identify different reasons for such failures. Researchers often cite inadequate funding, time constraints, company politics, technology limitations, and ethics obligations. While these factors can influence, they are more likely symptoms than causes. Users, on the other hand, often blame those who conducted the research. Their rationale is that the researchers are experts in their field and are thus culpable. If correct, research employees inside the company are dismissed and outside research firms are changed.

Responsibility and Involvement. Who is responsible for proper problem definition: users of the research or the researchers? Researchers are certainly not free of responsibility, since they are expected to discuss, explain, and clarify with the users, to identify precise and realistic goals. However, Exhibit 10.5 identifies four reasons for users to have greater responsibility.

Shifting Involvement in Problem Definition. While users are actually more involved than researchers at the beginning of problem definition, researchers very quickly become involved to a greater extent. Their involvement is especially important, since completing the problem definition serves as a formal agreement that provides general guidelines for the research project.

Three Dimension Pyramid. Problem definition is well represented by a three dimension pyramid, as Exhibit 10.6 illustrates. Each dimension is determined sequentially, beginning with the marketing research objective(s) and moving to the questions whose answers will meet the objectives and the information which when collected will answer the questions. While the sequence illustrated is directional, in reality there are substantial interdependencies.

Defining a marketing research problem means each dimension of the pyramid becomes a probe in the form of corresponding questions: (1) What are the objectives for conducting this marketing research project? (2) What are the questions to be answered by this marketing research project? (3) What information is expected to be provided by this marketing research project? Responses to each of these three questions must be realistic and mutually agreed upon by both the researcher and intended user of the research.

Explicit corresponding statements help to achieve realism and understanding. For example, statements at the end of the problem definition stage might be:

Exhibit 10.5
Reasons for Greater Users' Responsibility

(1) INITIATORS

Users initiate marketing research projects. As initiators, users know what is desired and they approach researchers who at this initial point have little or no idea about what a user wants.

(2) GREATER AWARENESS

By nature of their positions, users are closest to the marketing decisions to be made, and are thus more aware of the marketing decisions for which information is to assist.

(3) GREATER POWER

Users hold positions of greater power in the relationship. Since it is users who sponsor and fund particular marketing research projects, it can be interpersonally difficult for researchers to challenge the users' goals.

(4) GREATER STAKE

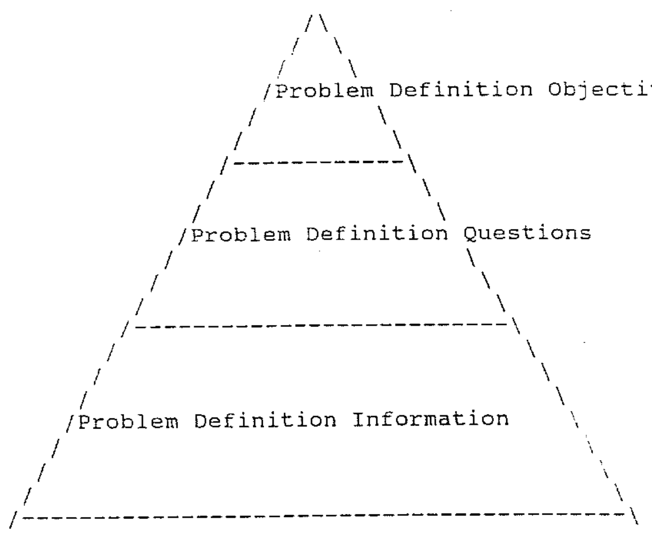
Users have more at stake. First, since it is users who fund marketing research projects, their money is expended whether or not a project attains its goals. Second, and more important, it is users who need the marketing research information to assist in decision making.

1. The objective(s) for conducting this research project is/are . . .
2. The question(s) to be answered by this marketing research project is/are . . .
3. The information expected to be provided by this marketing research project is . . .

Users benefit most by posing the above questions to themselves before meeting with a researcher. Then, during the initial meeting(s), users and researchers together determine the exact objective(s). In response, researchers formulate precise questions and specify appropriate information. During the final meeting, the researchers present explicitly the objectives, questions, and information pertinent to the marketing research project. When mutual understanding and agreement exists about these items, proceeding with the project is approved.

Explicitness is the most important rule of the format for problem definition. All objectives, questions, and information should be explicitly stated. While verbal statements are often sufficient, written statements are

Exhibit 10.6
Three-Dimension Pyramid Illustrating Problem Definition



preferred. Certainly, when a formal proposal is made, all of these statements are normally required in writing.

Three inherent concepts involve problem definition: objectives, questions, and information. Problem definition objective is the goal(s) that marketing research users have for a marketing research project. Before continuing the marketing research process, both the users and the researchers should be satisfied that the objectives are mutually understood, realistic, and measurable. Problem definition questions are questions whose answers collectively meet the problem definition objectives. Their collective answers provide information that meets the objectives that respective users have for a particular marketing research project. Problem definition information is information that is appropriate to collect to answer the specified problem definition questions. The ideal is to describe the information thoroughly and in detail.

Problems and Opportunities. Problem definition addresses opportunities as well as problems. Whether the situation confronted is negative or positive, problem definition involves specifying the marketing research objective(s) and questions. When marketing research addresses a negative situation, commonly referred to as a problem, it is important that the situation analysis differentiate between a symptom and a cause. Problem

Exhibit 10.7
Four Implicit Implications of Research Designs

Contractual relationship between researchers and users of research.
Road map and blueprint for performing marketing research project.
Transition between active involvement and little or no involvement by users of research.
Variation in format, detail, and formality.

symptom is the effect of a problem, and it is the result of a cause. Problem cause is the source of a problem, and it produces one or more symptoms.

Research Design

The research design is the specifications for conducting an entire marketing research project. Since it follows problem definition and precedes data collection,[2] its purpose is to carefully and thoroughly plan activities to achieve the objective(s) and answer the questions posed in the problem definition. A correspondingly detailed written document is highly desirable, but not always necessary.

It helps at the research design stage to keep in mind the standard list of questions—what, when, where, who, how, and why—for each of the components individually as well as collectively, for the overall marketing research project. The better these components are dealt with and these questions resolved, the better is the research design and ultimately the marketing research information provided to the decision makers.

The research design stage requires greater focus of the questions communicated earlier between users and researchers in the problem definition stage. The consequent is that these questions are now restated with emphasis on communication between researchers in the form of declarative statements (i.e., hypotheses). Plans at the research design stage include data type (e.g., secondary and primary, qualitative and quantitative, etc.), sample type (the people or entities to serve as subjects in the experiment), data collection (e.g., method, fieldwork, forms, mechanical and electronic devices, etc.), data analysis, report presentation (verbal and written reports), time schedule, and budget requirements.

Regardless of the exact components and their configuration, research designs have four implicit qualities (as listed in Exhibit 10.7). First, a research design establishes contractual implications between researchers and users of the research. It provides an agreement about what users expect researchers to do, and what researchers agree to do. If approved, the marketing research process continues. If not approved, the design is revised and again proposed. Second, it provides a "road map" or "blueprint" for researchers to follow. Third, it represents a transition between active involvement by users in the problem definition to little, if any, involvement in the remaining stages of the marketing research process.

Fourth, research designs vary substantially in format, detail, and formality. Depending on situation and user-researcher relationship, the range is from comprehensive written reports to mental ideas that are only vaguely discussed.

NOTES

1. While problem definition, versus opportunity definition, might be misleading for the uninformed, the phrase is technically correct. For example, dictionary definitions (*Chambers English Dictionary*, 7th ed., 1988, Cambridge, England: W. & R. Chambers Ltd. and Cambridge University Press, 1988) for the respective words are: Problem—"a question propounded for solution" (p. 1164); Define—"to fix the bounds or limits, to describe accurately" (p. 371); and Definition—"an explanation of the exact meaning" (p. 371). In summary, the users and doers of a marketing research project work together to put forward questions that request answers, and while putting forward these questions, the limits or boundaries are accurately described and exact meanings of the questions explained.
2. Different phrases are used in marketing and marketing research to refer to activities associated with the "research design." Two related phrases are "design the research" and "research design stage." All three of these phrases are used interchangeably, but to be precise, "design the research" is the action phrase referring to the activities associated with the research design stage of the marketing research process, "research design stage" is the group of activities necessary to design the research, and "research design" is the specifications for conducting an entire marketing research project.

Chapter 11

Two Levels of Quantitative Analysis: Percentages and Analysis of Variance

This chapter describes a complete marketing research experiment conducted by the author of this book. The parts characteristic of a complete experiment are included: background/context (often referred to as the literature review), methodology, results, and discussion. To illustrate different levels of quantitative analysis that can be performed with data collected through an experiment, results from two separate levels of data analysis are presented in two corresponding sections of the chapter. The first section demonstrates that data analysis limited to percentages can provide information useful to assist marketing decisions. The second section presents the results from performing analysis of variance of the data. Analysis of variance (commonly known as ANOVA) is a more sophisticated quantitative technique than percentages, and is frequently used to analyze data collected through experiments.

The experiment described was conducted to investigate the relative effectiveness of a 30-second television commercial versus a 15-second television commercial. A general reason for this experiment was that knowledge about the effectiveness of 15-second television commercials had not kept pace with marketing's increasing use of them.

SECTION ONE: AN EXPERIMENT WITH DATA ANALYSIS USING PERCENTAGES

BACKGROUND/CONTEXT

In North America, national television networks began offering 15-second commercials in the 1983–1984 television season. By 1987, 30-second commercials continued to dominate marketing strategies, but use of 15-second commercials had increased to 30% of all network advertising time and 20% of national advertising volume. By 1990, 15-second commercials comprised 40% of daytime commercials and industry executives were predicting increases to 50% daytime and over 40% prime time.

Early speculation was that 15-second commercials would soon dominate American television, but the peak appears to have occurred at 40% of television network spots.[1] Regardless, with nearly one-half of the television advertisements comprised of 15-second time periods, it was important for marketing managers to consider seriously their effectiveness as a tool within marketing strategy, especially relative to the traditional standard of the 30-second television commercial.

EFFECTIVENESS AND COST

The fact that knowledge had not kept pace with use meant that 15-second commercials were being used increasingly but their effectiveness was not documented. Early "research comment" suggested that effectiveness in terms of communication was 75% to 85% of 30-second commercials.[2,3] However, this difference was not identified through controlled experiments with external validity. Other research reported similar percents in the context of also exploring perceptions of the length of commercials and their impact on surrounding editorial context.[4]

Still another study concluded that recall of advertising copy was lower, on average, for 15-second commercials than for 30-second commercials.[5] The authors then stated that their study involved "exploratory, small-sample data" collected from less than 80 undergraduate college students. Their research must consequently be qualified for the marketing practitioner, since substantial research has emphasized the limitations for college students to serve as valid subjects in television commercial studies.[6]

Quite simply, students are rarely adequate subjects for an experiment when generalization of the findings is expected to represent an overall television audience. Students could possibly reflect sex differences analogous to people comprising the overall television audience, but the students would generally be too young and too well educated. Therefore, the experiment described in detail in this chapter involved "real-world" data collected from people outside a university setting.

In addition to the unanswered questions about relative effectiveness of 15- and 30-second television commercials, differences in the cost contributed to the need for a comprehensive experiment. For example, initially, the media cost for 15-second commercials was one-half the cost of 30-second commercials, but it shortly thereafter increased to more than 50% of the price for 30-second commercials, with local television stations charging up to 85% of the 30-second price for 15-second commercials. Part of

the reason for these disproportionate prices was a fear that 15-second commercials ultimately have a negative impact on television programs and advertising overall. As early as 1985, a report stated that "consumers describe [15-second commercials] less positively—as less interesting, less believable, less warm, less informative, more irritating and more confusing."[7] A year later, an industry executive stated that "overall, [15-second commercials have] been less effective than the [30-second commercial] and is responsible for increased negative viewer reaction, especially with the largest target group—the 18–34-year-olds."[8]

The reality was that difference in effectiveness may or may not be about one-fourth, while the cost differential was definitely no longer one-half. At the same time, decreasing network viewership and increasing advertising clutter raised questions about the impact of 15-second commercials on viewer satisfaction. For example, if 15-second commercials yield less program satisfaction, than their increasing use may be one possible factor contributing to declining viewership. A related consideration for marketing strategy is the possibility of promoting products in a program context where the consumers' state of mind is less satisfactory than an alternative program context.

In response, a formal experiment was conducted to assess relative effectiveness of 15- and 30-second commercials, with effectiveness operationally defined in terms of recall of brand and attitude toward brand. Additional measures assessed perception of commercial quantity and total commercial time, and program satisfaction.

METHODOLOGY

Television commercials were presented in the context of a standard network program. The program was a pilot for a situation comedy never before broadcast. A total of 12 30-second commercials (:30s) and 12 corresponding 15-second commercials (:15s) were used. All the :30s and :15s were submitted by corporate advertisers for network broadcast. The :15s were "lifts," identical to the :30s in theme, content, and style. All 12 brands were major national names.

Because of proprietary considerations, only the product categories are published in this article: (A) muffin mix, (B) window cleaner, (C) oral anesthetic for babies' teething pain, (D) mouth wash, (E) cough drops, (F) oven cleaner, (G) bar soap, (H) eye drops, (I) car wax, (J) pain killer for arthritis, (K) insect spray, and (L) cold medicine.

Three formats of the presentation, each with three one-minute commercial breaks, were used as illustrated in Exhibit 11.1. The result was that different groups of participants were exposed to different sets of commercials within the same program. To use a sample representative of the national American population, these people were recruited from tourist areas

Exhibit 11.1
Presentation Formats

FORMAT 1—15-SECOND COMMERCIALS ONLY

1st Break	2nd Break	3rd Break	(viewers)
A B C D	E F G H	I J K L	(185)

FORMAT 2—15-SEC. AND 30-SEC. COMMERCIALS

1st Break	2nd Break	3rd Break	
A B D*	E F* G	I* J K	(120)
C D A*	G H* F	K* L I	(120)
A C B*	E F* H	I* J L	(120)
B D C*	G H* E	K* L J	(120)

FORMAT 3—30-SECOND COMMERCIALS ONLY

1st Break	2nd Break	3rd Break	
A* B*	E* F*	I* J*	(180)
C* D*	G* H*	K* L*	(180)

Note: Each letter refers to a commercial for a different product. An asterisk (*) identifies a 30-second commercial.

in New York City and Los Angeles. Potential participants were approached by "recruiters" who were "blind" to the purpose of the experiment. To enhance external validity, data were collected through use of standard industry practices, facilities, and stimulus materials for marketing research related to national programming at one of the major American television networks.

Collectively, more than 6,000 recruited people viewed the different formats. Recruiters, unaware of the purpose of the experiment being conducted, approached approximately every twentieth adult on the street at different times during 15 different days (involving every day of the week). Each adult approached was given a standard verbal invitation to view a new television network program.

From these 6,000 people, 1,025 completed questionnaires were randomly selected for analysis. These questionnaires were selected through a two-step process that permitted matched demographic samples for each format. Without awareness of the questionnaire data, and within constraints of quotas, questionnaires were selected for analysis. The quotas reflected adult audience demographics consistent with the network's other, similar programs. Exhibit 11.2 presents the sample demographics.

Exhibit 11.2
Demographics and Configuration of the Sample

	Total Sample	Format 3 All :30s	Format 1 All :15s	Format 2 :30s & :15s
Sex				
Male	38%	38%	37%	37%
Female	62%	62%	63%	63%
Age				
18 - 34	37%	38%	39%	39%
35 - 49	23%	24%	23%	24%
50+	39%	39%	38%	37%
Education				
College	36%	37%	38%	39%
No College	64%	63%	62%	61%
Residence				
Metro NY or LA	33%	30%	33%	33%
Non-Metro NY/LA	67%	70%	67%	67%
Total Viewers	1,025	360	185	480

The entire pool of viewers, from which the sample was chosen, viewed the program in special viewing rooms in groups of about 25. Once seated, and before viewing the program, each viewer completed a brief questionnaire about his or her sex, age, education, occupation, and residence. All questionnaires were administered by a "session director." Completed questionnaires were delivered together by a session assistant, without comment, to a "sampler." The sampler used this information to select individuals consistent with the demographic quotas. Questionnaires from individuals associated with industries related to advertising, marketing research, or entertainment were eliminated.

Each viewer was given a posttreatment questionnaire after viewing the program. Types of questions included open-end essay, short-answer, and multiple response. The questionnaire asked viewers to evaluate the program and program elements, indicate the number of commercials and amount of commercial time, name the brand advertised in each product category (recall), and allocate points between brands (attitude). To measure recall, viewers were provided with the product category. To measure attitude, viewers were asked to allocate ten points between three competing brands listed for each product category (one of which was advertised).

Recall and attitude expressed by viewers who did not see a commercial for a particular brand served as the control. These control responses were

Exhibit 11.3
Brand Recall and Brand Attitude Across All Brands

	Saw :30	Saw :15	Did Not See
BRAND RECALL			
Percent Who Name Brand	20.9%	18.4%	8.9%
Increment	+12.0%	+ 9.5%	
BRAND ATTITUDE			
Average points given to brand	3.91	3.83	3.49
Increment	+.42	+.34	
TOTAL VIEWER MEASURES			
Brand Recall	3,600	5,100	3,600
Brand Attitude	3,600	5,100	3,600

used as a norm against which differences in brand recall and brand attitude were calculated. As Exhibit 11.1 illustrates, Format 2 viewers saw only 9 of the 12 test brands and Format 3 viewers saw only 6 of the 12 test brands.

RESULTS

Overall, :30s were approximately 20% more effective than :15s. Effectiveness was measured two ways: recall and attitude. Since viewers who saw only :30s or a mixture of :15s and :30s did not see commercials for all 12 products, their responses (regarding brands for which they did not see commercials) served as a no exposure/control group (i.e., norm).

Brand Recall

Brand recall was higher for viewers who saw a 30-second commercial than for those who saw a 15-second commercial. Exhibit 11.3 shows that, averaged across all brands and formats (i.e., the commercial may have been part of the mixed :15s and :30s format or with others of the same length), 20.9% of viewers named the advertised brand if they had seen a 30-second commercial for that brand. (Also, Exhibit 11.4 gives the percent who named the advertised brand correctly for each brand and each commercial format.) In contrast, 18.4% of viewers named the advertised brand if they had seen a 15-second commercial. Among the control group of viewers, 8.9% named the brand. The 20.9% for :30s is 12.0% above the 8.9% norm, and the 18.4% for the :15s is 9.5% above the norm. As a result, a single 15-second commercial, on average, was 79% as effective in terms of brand recall as a single 30-second commercial (9.5%/12.0% = 79.2%).

Exhibit 11.4
Percent Who Named the Advertised Brand Correctly, for Each Brand and Each Commercial Format (sample size is in parentheses)

Comm. Format: Length Seen:	Total Saw :15	All :15 Saw :15	:30 & :15 Saw :15	Total Saw :30	All :30 Saw :30	:30 & :15 Saw :30	Did Not See
(A) Comm. for muffin mix	22% (425)	20% (185)	23% (240)	23% (300)	23% (180)	23% (120)	11% (300)
(B) Comm. for window cleaner	11% (425)	12% (185)	9% (240)	15% (300)	16% (180)	14% (120)	3% (300)
(C) Comm. for oral anesthetic	23% (425)	23% (185)	22% (240)	25% (300)	24% (180)	25% (120)	2% (300)
(D) Comm. for mouth wash	11% (425)	11% (185)	11% (240)	9% (300)	11% (180)	7% (120)	1% (300)
(E) Comm. for cough drops	18% (425)	21% (185)	15% (240)	21% (300)	19% (180)	24% (120)	9% (300)
(F) Comm. for oven cleaner	34% (425)	35% (185)	33% (240)	38% (300)	36% (180)	40% (120)	24% (300)
(G) Comm. for bar soap	31% (425)	31% (185)	31% (240)	33% (300)	31% (180)	35% (120)	19% (300)
(H) Comm. for eye drops	34% (425)	31% (185)	36% (240)	39% (300)	39% (180)	38% (120)	23% (300)
(I) Comm. for car wax	9% (425)	7% (185)	11% (240)	14% (300)	16% (180)	12% (120)	2% (300)
(J) Comm. for arthritis pain	3% (425)	3% (185)	3% (240)	3% (300)	3% (180)	2% (120)	1% (300)
(K) Comm. for insect spray	17% (425)	17% (185)	16% (240)	21% (300)	20% (180)	21% (120)	9% (300)
(L) Comm. for cold medicine	8% (425)	9% (185)	7% (240)	10% (300)	11% (180)	9% (120)	3% (300)

Brand Attitude

Brand attitude was more favorable for viewers who saw a 30-second commercial than for those who saw a 15-second commercial. Averaged across all brands, Exhibit 11.3 shows that viewers gave an average of 3.91 points to an advertised brand if they had seen its 30-second commercial and 3.83 points if they had seen its 15-second commercial. (Also, Exhibit 11.5 presents the average points given to the advertised brand, for each brand and each commercial format.) Among the control group, an average

Exhibit 11.5
Average Points Given to the Advertised Brand, for Each Brand and Each Commercial Format (sample size is in parentheses)

Comm. Format: Length Seen:	Total Saw :15	All :15 Saw :15	:30 & :15 Saw :15	Total Saw :30	All :30 Saw :30	:30 & :15 Saw :30	Did Not See
(A) Comm. for muffin mix	4.24 (425)	4.33 (185)	4.16 (240)	4.29 (300)	4.30 (180)	4.28 (120)	3.97 (300)
(B) Comm. for window cleaner	3.10 (425)	3.14 (185)	3.06 (240)	3.23 (300)	3.28 (180)	3.14 (120)	2.72 (300)
(C) Comm. for oral anesthetic	4.86 (425)	4.78 (185)	4.94 (240)	5.18 (300)	5.29 (180)	4.95 (120)	4.13 (300)
(D) Comm. for mouth wash	2.72 (425)	2.89 (185)	2.57 (240)	2.74 (300)	2.79 (180)	2.64 (120)	2.70 (300)
(E) Comm. for cough drops	4.02 (425)	4.30 (185)	3.75 (240)	3.94 (300)	3.87 (180)	4.07 (120)	3.61 (300)
(F) Comm. for oven cleaner	4.64 (425)	4.44 (185)	4.83 (240)	4.67 (300)	4.72 (180)	4.57 (120)	4.08 (300)
(G) Comm. for bar soap	4.65 (425)	5.07 (185)	4.24 (240)	4.81 (300)	4.74 (180)	4.96 (120)	4.32 (300)
(H) Comm. for eye drops	5.62 (425)	5.77 (185)	5.48 (240)	5.38 (300)	5.40 (180)	5.33 (120)	5.15 (300)
(I) Comm. for car wax	2.44 (425)	2.29 (185)	2.59 (240)	2.74 (300)	2.84 (180)	2.51 (120)	1.94 (300)
(J) Comm. for arthritis pain	2.85 (425)	2.86 (185)	2.83 (240)	2.81 (300)	2.73 (180)	2.97 (120)	2.59 (300)
(K) Comm. for insect spray	3.36 (425)	3.32 (185)	3.39 (240)	3.67 (300)	3.61 (180)	3.80 (120)	3.35 (300)
(L) Comm. for cold medicine	3.48 (425)	3.43 (185)	3.53 (240)	3.46 (300)	3.44 (180)	3.49 (120)	3.31 (300)

of 3.49 points were allocated to the brand. The 3.91 score for :30s is +.42 points higher than the 3.49 norm. The 3.83 score for :15s is +.34 points higher than the norm. As a result, a single 15-second commercial, on average, was 81% as effective in terms of brand attitude as a single 30-second commercial (.34/.42 = 81.0%).

Commercial Units

As the number of commercials increased during the 30-minute program, the number perceived also increased. As Exhibit 11.6 shows, each actual

Exhibit 11.6
Perceived Quantity and Total Time of Commercials

	Format 3 Saw Only :30	Format 2 Saw :30 & :15	Format 1 Saw Only :15
PERCEIVED QUANTITY OF COMMERCIALS			
Average # in program	5.5	7.9	8.6
Percent difference (base: actual)	-0.08% (6 actual)	-12.2% (9 actual)	-28.3% (12 actual)
Relative # estimated			
More than usual	12%	20%	27%
Same as usual	58%	59%	59%
Fewer than usual	30%	19%	14%
PERCEIVED TOTAL TIME DEVOTED TO COMMERCIALS			
Average length of time	5.2 mins.	6.6 mins.	5.6 mins.
Percent difference (base: actual)	+73.3% (3.0 actual)	+120.0% (3.0 actual)	+86.7% (3.0 actual)
Relative time estimated			
More time than usual	10%	14%	20%
Same time as usual	54%	59%	52%
Less time than usual	37%	26%	26%
TOTAL VIEWERS			
Perceived Quantity	360	480	185
Perceived Total Time	360	480	185

increase caused viewers to perceive a greater number of commercials. These numbers ranged from 5.5 (based on the Format 3 presentation with six commercials) to 7.9 (based on Format 2 with nine commercials) to 8.6 (based on Format 1 with twelve commercials). However, estimates were less than actual for all three formats: .08% less based on the :30s format, 12% less based on the mixed format, and 28% less based on the :15s format.

A similar pattern was revealed for the *relative* number of commercials. In relative terms, as the number of commercials increased from format to format, larger proportions of viewers felt there were more commercials than usual (12% based on all :30s, 20% based on mixed lengths, and 27%

based on :15s). Regardless of which format viewers were exposed to, the majority of viewers (58% to 59%) thought there were the same number as usual. Considering commercials normally at the beginning and end of programs, promotional spots, public service announcements, and station identifications (none of which were part of this test), there were actually fewer "commercials" than usual.

Commercial Time

In all three presentation formats, there were three minutes of commercial time. But as Exhibit 11.6 shows, all groups of the viewers perceived substantially more time devoted to commercials than there actually was. Also, even though all the format times were equal, viewers who saw more commercials perceived that there was more time devoted to commercials. These perceptions were 5.2 minutes based on all :30s, 6.6 minutes based on the mixed format, and 5.6 minutes based on all :15s. The difference between estimated time and actual time ranged from 120% more for mixed format, 87% more based on all :15s, and 73% more based on all :30s.

Again, a similar pattern was revealed for the *relative* time devoted to commercials. In relative terms, as the number of commercials increased from format to format, larger proportions of viewers felt there was "more time devoted to commercials than usual" (10% based on all :30s, 14% based on mixed lengths, and 20% based on :15s). Regardless of which format viewers were exposed to, the majority of them (52% to 59%) thought there was the same time as usual.

Program Evaluation

Exhibit 11.7 shows there were not substantial differences between how the viewers of the three commercial formats rated the program. Nearly one-half of all the viewers (i.e., those who saw only :15s, those who saw a mixture of :15s and :30s, and those who saw only :30s) said they would watch future episodes of the program if the episodes were very much like the one they saw. Nearly one-fourth of each group said the program was better than most, while well over one-half of each group said they would recommend the program to friends.

DISCUSSION

The results of this experiment yield managerial implications and recommendations. Despite the procedural limitation of a single exposure to multiple commercials, these results suggest that a reasonable conclusion is that 30-second commercials are more effective than 15-second commercials. The implication is that a marketing manager is likely to achieve

Exhibit 11.7
Program Evaluations

	Saw Only :30	Saw :30 & :15	Saw Only :15
WOULD WATCH FUTURE EPISODES IF:			
Very much like this one	45%	47%	45%
Improved a bit	30%	31%	29%
Improved a good deal	15%	14%	19%
Would never turn on	10%	8%	7%
PROGRAM IS:			
Better than most	23%	20%	23%
About as good as most	58%	56%	58%
Not as good as most	19%	23%	19%
WOULD RECOMMEND TO FRIENDS:			
Yes	64%	63%	65%
No	36%	36%	34%
TOTAL VIEWERS FOR EACH QUESTION ITEM:			
Would watch...	360	480	185
Program is...	360	480	185
Would recommend...	360	480	185

greater brand recall and more favorable brand attitude with 30-second commercials than with 15-second commercials. The differential effectiveness is approximately 20% less for 15-second commercials compared to 30-second commercials.

The 20% difference in effectiveness should be a definite consideration in allocating marketing budgets. If a marketing manager is confronted with media costs for 15-second time slots that are greater than 80% of the media costs for 30-second time slots, it is advantageous to allocate budget expenditures in favor of 30-second commercials. On the other hand, if media costs for 15-second time slots are less than 80% of the media costs for 30-second time slots, it is advantageous to allocate budget expenditures in favor of 15-second commercials. The general rule should be:

IF :15s cost > 80% of :30s cost THEN favor :30s
IF :15s cost < 80% of :30s cost THEN favor :15s

When purchasing media time in the past, this 80% point was typically not very serious. But the past 50% price difference no longer prevails. Even as early as 1987, some local television stations already were charging 85%

of the :30s price.[9] Therefore, in the future, this 80% effectiveness difference is likely to become an increasingly more serious consideration.

State of Mind

A relevant factor to consider in budget allocation decisions is the state of mind of consumers during 30-second and 15-second commercials. One aspect of this consideration is the "editorial context" in the form of surrounding program.

A consumer's state of mind while watching a program is not likely to be divorced from the attendant commercials. It is reasonable to speculate that state of mind during the program will generalize and be similar during the interspersed commercials. Factors that impact feelings about a program likely impact state of mind during the program, and ultimately during the commercials. Given the general dislike of commercials by consumers, the increased number of commercials associated with 15-second commercials becomes one such factor.

Intuition suggests that 15-second commercials induce a correspondingly less favorable state of mind among viewers. The impact is ultimately less positive evaluation of the program caused by the perception of more commercials and more commercial time. While intuitively appealing, this speculation is not supported by the results of this experiment.

Replacing 30-second commercials with 15-second commercials did cause viewers to perceive more commercials and more total commercial time, in both absolute and relative terms. However, the results revealed that regardless of whether the program was viewed in a setting of 15-second commercials, 30-second commercials, or a mixture of 15-second and 30-second commercials, the different formats did not alter satisfaction with the program, intent to watch future episodes, or intent to recommend the program to friends. The implication for marketing managers is that imbedding 15-second commercials in television programs does not adversely impact the consumer's evaluation of the program (i.e., the editorial context).

Future Research

The current experiment attempted to reflect realities of the marketplace (i.e., external validity within the confines of true cause-and-effect inquiry). Its results add to the overall knowledge of :15s by providing particular insight into the prudence of using marketing strategies that employ 15-second commercials rather than 30-second. At the same time, many aspects of this current experiment raise the need for future research to address many related dimensions and questions that remain to be identified, investigated, and answered.

In regard to commercials, while viewers were presented with multiple

commercials during the 30-minute program, each commercial was shown only once. Future experiments should question if the differences identified in this situation exist in other situations. It is well-known that the consequence of a unitary dose of a commercial is different than the consequence of repeated exposures to the same commercial, but is the consequence the same for multiple exposures to the same 15-second commercial and the same 30-second commercial?

In regard to programs, a 30-minute program was used. Experiments in the future should investigate if the impact is different for :15s and :30s when shown in the context of different program lengths (e.g., 30 minutes, 60 minutes, two hours, an entire evening, consecutive evenings with a miniseries, etc.). Do different program genres (e.g., situation comedy, action-adventure, news, etc.) impact :15s and :30s differently? Are evaluations of the programs themselves affected differently (when presented with 15-second and 30-second commercials) during the different options for program length? The program longevity may also yield different effects. Are the consequences different for a long-run, familiar program and a new, never-seen program?

There are also product life cycle considerations. Are impact differences between :15s and :30s the same for new, less familiar products and not new, well-established products?

The global marketplace provides another dimension for research to address and for marketing managers to react. In fact, the use of :15s in other countries may foretell America's future. If the past increases of :15s persist in the United States, the American marketplace could soon equal the frequency, and possibly the acceptance, exhibited in other countries. Already, when use of :15s was just beginning in the United States, reports were that "[i]n Japan, 15-second spots are a part of the commercial culture,"[10] with 70% of all commercials in Japan at that time already less than 30 seconds.[11] As business continues to become more international, future research providing such comparative information holds crucial implications for successful marketing strategies.

SUMMARY/ABSTRACT

An experiment was used to investigate differences in effectiveness between 30-second and 15-second television commercials. Subjects were adults from across the United States. As well as measures of effectiveness (recall and attitude), commercial perceptions (quantity and time in both absolute and relative terms) and program evaluations (to watch future episodes, in comparison to other programs, and to recommend to friends) were assessed.

Results showed 15-second commercials were about 20% less effective than 30-second commercials. This 20% was the average across 12 product

categories in which there was variation from 42% less effective to 25% more effective. Results also showed that imbedding a program with 15-second commercials did not alter viewers' satisfaction with the program. But use of 15-second commercials did cause perception of more commercials and more commercial time, in both absolute and relative terms.

The 80% relative effectiveness shown in this experiment suggests a valuable factor to consider in marketing strategy plans. It may be advantageous to allocate budget expenditures in favor of 15-second commercials as long as the media costs are less than 80% of the media costs for 30-second time slots. It is no longer advantageous to make allocations that favor 15-second commercials when the cost of 15-second time slots exceeds 80% of the cost of 30-second time slots. This consideration about relative effectiveness versus relative price is especially valuable given the disproportionate price increases that have occurred for 15-second television commercials.

SECTION TWO: DATA ANALYSIS USING ANALYSIS OF VARIANCE

While percentages can be useful, and at times adequate, to assist marketing decision makers, data from an experiment are often analyzed with additional quantitative techniques. One frequently used technique is analysis of variance as described in this section. Since these data were collected through the same experiment described above, description of the background and methodology are not repeated here. Most aspects of the design, such as stimulus materials, subjects, and stimulus presentation format, were also the same and are also omitted here. However, additional details about the design are presented as they pertain first, to commercial quantity and commercial time, and second, to editorial context.

COMMERCIAL QUANTITY AND COMMERCIAL TIME

The motivation for this experiment and these data analyses was that despite knowledge that consumers generally react negatively to more commercials and more commercial time, research about consumers' respective perceptions has not paralleled the increase in the use of 15-second commercials.

Design

The research design conducted to investigate these perceptions was a $2 \times 3 \times 3 \times 2$ factorial experiment. These factors were one independent treatment variable and three independent blocking variables. The independent treatment variable was television commercial length (15-second commercials and 30-second commercials). The three independent blocking variables were age (18–34, 35–49, and 50+), education (college, high school, and grammar school), and sex (male and female).

The dependent variables were posttreatment measures of the perceived: absolute quantity of commercials, quantity of commercials relative to usual, absolute total time devoted to commercials, and total time devoted to commercials relative to usual. Collectively, more than 2,500 people viewed the different commercial formats and completed questionnaires. From these viewers, 655 subjects were selected randomly to comprise the sample. No subjects participated in repeated treatment measures.

Sample subjects were selected randomly through a blind, two-stage process that involved two separate questionnaires to select representative demographics without awareness of their questionnaire data. The random selection was guided by an objective of matched samples for each format, eight subjects maximum from any one viewing session, and adherence to demographic quotas. These quotas reflected adult audience demographics consistent with the network's other similar prime-time programs.

Commercials versus Programs

Assessing perceived quantity and total time of television commercials is not necessarily the same as testing television programs. The testing of television programs focuses on attitudes, but here the focus was on perceptions. This difference poses a valid concern. First, it is likely that consumers devote little mental energy to television advertising when in their home living room. However, a laboratory experiment of advertising may tend to focus untypical attention on the stimuli. Second, it is reasonable to speculate that more mental energy is necessary to recall perceptions than to express attitudes.

Hypotheses

Two hypotheses were proposed, each with two parts.

Hypothesis 1-A: Subjects (comprised of typical television audience members) exposed to 15-second commercials will perceive a greater quantity of commercials than subjects exposed to the same commercial time comprised of 30-second commercials.

Hypothesis 1-B: A larger proportion of these subjects will perceive more commercials relative to usual when exposed to the same commercial time comprised of 15-second commercials compared to 30-second commercials.

Perceptual acuity of the typical American adult television audience member is likely able to discriminate between :15s and :30s. This acuity is especially likely when commercials are of a sufficiently low quantity, during the standard one-half-hour, prime-time, national television network program, that television audiences can probably sense a substantial change. Commercial breaks comprised solely of :15s present a 100% increase of intensity (in terms of quantity) over commercial breaks comprised solely of :30s. It is reasonable to expect that television audience members exposed to this higher quantity of commercials perceive more commercials.

There are two reasons. First, repeated past exposures have acclimated audience members to a lower quantity intensity (i.e., fewer separate commercial messages in the case of :30s). Second, a 100% change from the prior low intensity is likely to exceed values for JND (just noticeable difference) thresholds. Therefore, part A of this first hypothesis would be rejected if subjects exposed to :15s perceive an equal or smaller quantity of commercial units than subjects exposed to :30s. Part B of this first hypothesis would be rejected if the proportion of subjects exposed to :15s, that indicate "more commercials relative to usual," is equal to or less than the corresponding proportion of subjects exposed to :30s.

Hypothesis 2-A: Subjects (comprised of typical television audience members) exposed to 15-second commercials will perceive a greater amount of total time devoted to commercials than subjects exposed to the same commercial time comprised of 30-second commercials.

Hypothesis 2-B: A larger proportion of these subjects will perceive more total time devoted to commercials relative to usual when exposed to the same commercial time comprised of 15-second commercials compared to 30-second commercials.

Early industry data revealed that "most viewers cannot tell the difference in length between a 30-second commercial and a 15-second commercial."[12] These data provide a priori support for this hypothesis. Hypothesis 2 is also supported by simple multiplication; if Hypothesis 1 fails to be rejected. For example, subjects are exposed to four :15s versus two :30s with each one-minute commercial break, and to twelve :15s versus six :30s with the entire one-half-hour program. Even though the length of time is the same, subjects are expected to perceive that more total time is being devoted to commercials when more separate commercials are presented. Therefore, part A of this second hypothesis would be rejected if subjects exposed to :15s perceive equal or less total time devoted to commercials than subjects exposed to :30s. Part B of this second hypothesis would be rejected if the proportion of subjects exposed to :15s, that indicate "more total time devoted to commercials than usual," is equal to or less than the corresponding proportion of subjects exposed to :30s.

As a footnote, the movie industry provides a different perspective for this hypothesized perception of greater total commercial time. Directors have long used a prolonging technique by changing camera angles while focused on a single scene. These camera angle changes disrupt the movie viewers' focus, which in turn prolongs the perception of time that a single scene is shown. An illustration of this technique and effect is the "shower scene" in the classic movie *Psycho* by director Alfred Hitchcock. Similar to this moviemaking technique, a one-minute commercial break with four :15s (versus two :30s) presents more camera angle changes, which through analogy offers an explanation for the perception of greater total time for :15s compared to :30s.

RESULTS

Measures of the dependent variables were treated as interval data for analyses. These data involved semantic differential scales to measure relative quantity and time, and open-end questions to collect numerical estimates about absolute quantity and time.

Manipulation Check

Manipulation of the independent treatment variable was successful. More commercials were perceived in the :15s condition than the :30s condition. In response to an open-end question to estimate the number of commercials during the program, the mean number for the :15s treatment was 10.66 and for the :30s it was 5.63. Analysis of variance (with perceived number of commercials as the criterion) showed a significant main effect for commercial length ($F = 690.62$, df = 1,654, $p < .001$). It may be worthwhile to again note there were no repeated treatment measures.

Hypothesis 1

Hypothesis 1 was not rejected. As hypothesized in 1-A, subjects perceived a greater quantity of commercials when exposed to :15s than :30s. Also as hypothesized in 1-B, a larger proportion of subjects perceived more commercials relative to usual when exposed to :15s than :30s.

Quantity of commercials was measured by asking an open-end question: "About how many commercials would you say were in the program?" Relative quantity was measured by asking: "Would you say there were more commercials, fewer commercials, or about the same number of commercials as are usually included in a half-hour television show?" The labelled choices were: fewer, same, and more.

Hypothesis 1-A. Part A of hypothesis 1 was not rejected. Subjects exposed to :15s perceived a greater quantity of commercials than subjects exposed to the same commercial break time comprised of :30s. Analysis of variance with perceived number of commercials as the criterion showed a significant main effect for commercial length (F = 690.62, df = 1,654, $p < .001$). Mean number of commercials perceived was 10.66 for subjects exposed to :15s and 5.63 for subjects exposed to :30s. A significant main effect was not shown for age (F = 1.02, df = 2,654, $p < .363$), but was shown for education (F = 24.87, df = 2,654, $p < .001$) and sex (F = 8.01, df = 1,654, $p < .005$). In terms of education, the mean number of commercials perceived was 7.01 (college), 7.62 (high school), and 9.27 (grammar school). In terms of sex, the mean number of commercials perceived was 7.20 (males) and 7.82 (females).

A significant interaction did not occur between commercial length and age (F = 1.95, df = 2,654, $p < .143$) or between commercial length and sex (F = 0.01, df = 1,654, $p < .918$), but did occur between commercial length and education (F = 13.02, df = 2,654, $p < .001$). For both :15s and :30s, college- and high-school-educated subjects perceived fewer commercials than the actual quantity (ranging from 5% less to 19% less), while grammar-school-educated subjects perceived 8% to 13% more commercials than the actual quantity (Exhibit 11.8). For all levels of education, the amount of difference between perceived quantity and actual quantity was greatest for :15s: −19% versus −13%, −11% versus −5%, and +13% versus +8%.

Hypothesis 1-B. Part B of hypothesis 1 was not rejected. Subjects exposed to :15s perceived a greater quantity of commercials relative to usual than subjects exposed to :30s. For purposes of analysis, responses to the question about quantity of commercials relative to usual were assigned the values of fewer, same, and more. Subjects were then categorized into respective groups.

Analysis of variance showed a significant main effect for commercial length (F = 14.24, df = 1,654, $p < .001$). Mean values for relative quantity of commercials were 2.05 (or 2.5% more than the same-as-usual value) for subjects exposed to :15s and 1.89 (or 5.5% less than the same-as-usual value) for subjects exposed to :30s. A significant main effect was also shown for age (F = 5.04, df = 2,654, $p < .007$), education (F = 5.51, df = 2,654, $p < .004$), and sex (F = 13.60, df = 1,654, $p < .001$). Exhibit 11.9 presents the mean values.

The proportion of subjects (Exhibit 11.10) perceiving more commercials than usual was twice as high when exposed to :15s (22.0%) as :30s (11.8%). Regardless of whether subjects were exposed to :15s or :30s, the overwhelming majority (78.1% based on :15s and 88.3% based on :30s) did not perceive that the quantity of commercials was more than usual.

A significant interaction did not occur between commercial length and

Exhibit 11.10
Proportions of Subjects Regarding Perceived Quantity of Commercials Relative to Usual Based on Commercial Length

Relative to Usual	Comm. Length :15s	Comm. Length :30s	Difference :30s-:15s
Fewer	16.9%	23.0%	+6.1 percentage points
Same	61.2%	65.3%	+4.1 percentage points
More	22.0%	11.8%	-10.2 percentage points

df = 1,654, $p < .135$), but was shown for education (F = 26.32, df = 2,654, $p < .001$). In terms of education, the mean total commercial time perceived was 5.45 minutes (college), 6.27 minutes (high school), and 7.92 minutes (grammar school).

No significant interactions occurred between commercial length and age (F = 0.06, df = 2,654, $p < .946$) or commercial length and education (F = 1.58, df = 2,654, $p < .207$), but a significant interaction did occur between commercial length and sex (F = 20.03, df = 1,654, $p < .001$). Both groups perceived more total time devoted to commercials when the program was presented with :15s rather than :30s (Exhibit 11.12). For male subjects, it was an additional 2.91 minutes, while for female subjects it was an additional 1.01 minutes.

Hypothesis 2-B. Part B of hypothesis 2 was not rejected. Subjects exposed to :15s perceived that greater total time was devoted to commercials relative to usual than subjects exposed to :30s. For purposes of analysis, responses to the question about total time devoted to commercials relative to usual were assigned the values of less, same, and more. Subjects were then categorized into respective groups based on these data.

Analysis of variance showed a significant main effect for commercial length (F = 20.74, df = 1,654, $p < .001$). Mean values for relative total commercial time were 1.95 (or 2.5% less than the same-as-usual value) for subjects exposed to :15s and 1.73 (or 13.5% less than the same-as-usual value) for subjects exposed to :30s. A significant main effect also was revealed for age (F = 4.93, df = 2,654, $p < .008$) and education (F = 5.11, df = 2,654, $p < .006$), but not for sex (F = 1.63, df = 1,654, $p < .202$). The mean values for these respective cells are presented in Exhibit 11.13.

The proportion of subjects (Exhibit 11.14) perceiving more total time devoted to commercials than usual was higher when exposed to :15s (16.4%) than :30s (10.3%). However, regardless of whether subjects were

Exhibit 11.11
Significant ANOVA Interaction Effects: Perceived Quantity of Commercials Relative to Usual

	:15s**	%Same	:30s**	%Same
Education***				
College	2.14	+7%	1.72	-14%
High School	2.01	+0.5%	1.92	-4%
Grammar School	1.97	-1.5%	2.27	+13.5%

*Relative-to-Usual values were: Fewer=1, Same=2, and More=3.

**Mean values

***Significant ANOVA interaction with commercial length ($p<.001$).

Note: Only these means are presented because ANOVA showed a significant interaction between commercial length and level of education; and no significant interaction between commercial length and age ($p<.059$) or commercial length and sex ($p<.598$).

exposed to :15s or :30s, the overwhelming majority (83.6% based on :15s and 89.8% based on :30s) did not perceive that the total time devoted to commercials was more than usual. This overwhelming majority supports the early report that "most viewers cannot tell the difference in length" between :15s and :30s.[13]

A significant interaction occurred between both commercial length and age ($F = 5.74$, df = 2,654, $p < .003$) and commercial length and education ($F = 13.36$, df = 2,654, $p < .001$), but did not occur between commercial length and sex ($F = 1.02$, df = 1,654, $p < .313$). In terms of age, the mean value for the perception of more total time devoted to commercials relative to usual (Exhibit 11.15) was largest among young subjects exposed to :15s (2.10). The largest mean value difference from same as usual was among the young subjects exposed to the :30s (−20%).

In terms of education, the mean value for perception of more total time devoted to commercials relative to usual (Exhibit 11.15) was slightly above the same as usual for only college-educated subjects exposed to :15s (2.14). For all other education levels, mean values indicated perceptions were near or below the same-as-usual total time devoted to commercials, regardless of exposure to :15s or :30s.

Exhibit 11.12
Significant ANOVA Interaction Effects: Perceived Total Time Devoted to Commercials

	Comm. Length* :15s	Comm. Length* :30s	Difference :15s-:30s
Sex**			
Male	7.79	4.88	+2.91 mins.
Female	6.88	5.87	+1.01 mins.

*Mean values (in minutes)

**Significant ANOVA interaction with commercial length ($p<.001$)

Note: Only these means are presented because ANOVA showed a significant interaction between commercial length and sex; and no significant interaction between commercial length and age ($p<.946$) or commercial length and education ($p<.207$).

SUMMARY/ABSTRACT

Impact of commercial length on perception of commercial quantity and commercial time was the focus of this experiment. The methodology involved an experiment with a national sample of 655 American adults. Collected data assessed impact of 15-second commercials (:15s) and 30-second commercials (:30s). Four measures of impact on perception were used: absolute number of commercials, relative number of commercials, absolute total commercial time, and relative total commercial time.

A larger quantity of commercials, in both absolute terms and relative terms, was perceived when a television program was presented with :15s instead of :30s. A larger amount of total commercial time, in both absolute terms and relative terms, was also perceived based on :15s. Compared to their respective counterparts, these perceptions based on :15s relative to :30s were consistently highest for the less educated, tended to be higher for males, and were mixed for groups according to age.

EDITORIAL CONTEXT

The motivation for this experiment and these data analyses was a fear of 15-second commercials exhibited by the marketing industry. The fear is not related to effectiveness of :15s, but to the popularity of :15s. The concern is that :15s will translate into increased commercial clutter that ultimately causes less effective advertising. Based on this assumed causal link

Exhibit 11.13
Significant ANOVA Main Effects: Perceived Total Time Devoted to Commercials Relative to Usual According to Demographics

	Mean Values	%Difference From "Same"
Age**		
18-34	1.80	-10.0%
35-49	1.92	- 4.0%
50+	1.77	-11.5%
Education***		
College	1.77	-11.5%
High School	1.81	- 9.5%
Grammar School	1.96	- 2.0%
Sex****		
Male	1.85	+ 7.5%
Female	1.80	-10.0%

*Relative-to-Usual values were: Less=1, Same=2, and More=3.

**Significant ANOVA interaction with commercial length ($p<.008$).

***Significant ANOVA interaction with commercial length ($p<.006$).

****Non-significant ANOVA interaction with commercial length ($p<.202$).

Note: Means are presented for age, education, and sex, but ANOVA did not show a significant main effect for sex.

to clutter, the marketing industry has resisted use of :15s from the beginning, because "as far as many advertisers were concerned, they were simply fearful that increased clutter would lead to less attention paid to their commercials."[14]

Early research comment fostered these fears, since consumers were reported to describe :15s less positively as less interesting, less believable, less warm, less informative, more irritating, and more confusing.[15] A year later an industry executive stated that "overall, it [use of :15s] has been less effective than the :30 and is responsible for increased negative viewer reaction, especially with the largest target group—the 18–34-year-olds."[16] Research does show individual :15s, compared to :30s, are less effective

Exhibit 11.14
Proportions of Subjects Regarding Perceived Total Time Devoted to Commercials Relative to Usual Based on Commercial Length

Relative to Usual	Comm. Length :15s	:30s	Difference :30s-:15s
Less	21.6%	37.0%	+15.4 percentage points
Same	62.0%	52.8%	-9.2 percentage points
More	16.4%	10.3%	-6.1 percentage points

and evoke less favorable reactions, but existing research does not show that :15s have caused, or will cause, an actual measurable decrease in overall commercial effectiveness (through clutter or some other process).

Questions about the relationship between use of :15s and possible reduced overall commercial effectiveness reflect valid concerns. Associated clutter may lead to mental confusion about the commercials and/or less favorable reactions. A related concern is that associated clutter may ultimately lead to viewing behavior responses to avoid entire commercial breaks (either through remote control "edits" or nonselection of particular programs). Since editorial context is an integral part of commercial delivery, neither can be disregarded by the marketing industry. The fear remains, despite lack of research knowledge to justify it. Therefore, since these views are "not based on scientific research,"[17] the current experiment was conducted to investigate the impact of :15s on consumers' reactions to the editorial context, that is, the programming.

Marketers must be concerned about the relative effectiveness of :15s compared to :30s. However, even before the effectiveness stage, marketers should be concerned that consumers are at least physically exposed to their commercials and, it is hoped, while in a favorable state of mind. Therefore, viewers' reactions to the editorial context (i.e., the attending television program itself) are consequential.

Design

The design, a $2 \times 2 \times 3 \times 3$ factorial experiment, was the same as described earlier in this chapter. Likewise, most aspects of the design, such as stimulus materials, subjects, and stimulus presentation format, were the same as described above. Dependent variables focused on the program: interest to watch future episodes, curiosity to see the ending, comparison to similar programs, and likelihood of program recommendation to friends.

Exhibit 11.15
Significant ANOVA Interaction Effects: Perceived Total Time Devoted to Commercials Relative to Usual by Format

	:15s**	%Same	:30s**	%Same
Age***				
18-34	2.10	+ 5%	1.61	-20%
35-49	1.98	- 1%	1.88	- 6%
50+	1.77	-12%	1.77	-12%
Education****				
College	2.14	+ 7%	1.54	-23%
High School	1.80	-10%	1.82	- 9%
Grammar School	1.97	- 2%	1.96	- 2%

*Relative-to-Usual values were: Less=1, Same=2, and More=3.

**Mean values

***Significant ANOVA interaction with commercial length ($p<.003$).

****Significant ANOVA interaction with commercial length ($p<.001$).

Note: Only these means are presented because ANOVA showed a significant interaction between commercial length and age and between commercial length and level of education; and no significant interaction between commercial length and sex ($p<.313$).

These dependent variables were measured through a posttreatment questionnaire with semantic differential scales.

Hypotheses

The hypotheses of this experiment were based on existing published research. Extending beyond this base, it was speculated that television viewers make less-than-complete discrimination between commercials and programs. Specifically, consumers generalize their reactions during commercial breaks in a manner that carries over to be similar during attending programs, and vice versa. These hypotheses were consistent with generally negative reports about :15s relative to :30s.

Regardless of whether these hypotheses were rejected or failed to be rejected, they carried implications for marketing strategy. For example, if television viewers are not gratified by watching a particular program, this lack of gratification is likely to translate into a less favorable state of mind during the program and during the subsequent commercials. Or maybe even worse, it might translate into potential viewers who simply do not tune in the program at all (nor recommend that their friends do so).

The result for marketing managers is that programs presented with :15s could be a distinctly less desirable delivery vehicle. These less gratified program viewers either will be exposed to a commercial in a less favorable state of mind, or they (and their friends) will not even have the opportunity to be exposed. Ultimately, with either scenario, the initial negative impact of the :15s on reactions to a program will in turn produce a negative impact on the effectiveness of the commercials.

> Hypothesis 1: Interest to watch future episodes of a television program series will be lower when the program is presented with 15-second commercials than when presented with 30-second commercials, among typical television audience members.

This first hypothesis would be rejected if the interest to watch future episodes is equal or stronger among viewers who see the program presented with :15s compared to viewers who see the program presented with :30s.

> Hypothesis 2: Curiosity about the ending of an episode will be lower when the program is presented with 15-second commercials than when presented with 30-second commercials, among typical television audience members.

This second hypothesis would be rejected if the curiosity to see the ending of the episode is equal or greater among viewers who see the program presented with :15s compared to viewers who see the program presented with :30s.

> Hypothesis 3: Comparisons of a program with other similar programs will be less favorable when the program is presented with 15-second commercials than when presented with 30-second commercials, among typical television audience members.

This third hypothesis would be rejected if the comparison of the program is equal or more favorable among viewers who see the program presented with :15s compared to viewers who see the program presented with :30s.

> Hypothesis 4: Recommendation of a program to friends will be less likely when the program is presented with 15-second commercials than when presented with 30-second commercials, among typical television audience members.

This fourth hypothesis would be rejected if the likelihood of recommendation, of the program to friends, is equal or more likely among viewers who see the program presented with :15s compared to viewers who see the program presented with :30s.

RESULTS

Hypothesis 1

Hypothesis 1 was rejected. Different than hypothesized, interest to watch future episodes was not significantly lower when the program was presented with :15s compared to :30s. Analysis of variance did not show a significant main effect for commercial length ($F = .759$, df = 1,654, $p < .759$). Mean values for interest to watch future episodes of this program were 3.09 for :30s viewers and 3.12 for :15s viewers. Among this national sample of typical adult television audience members, relative interest as a function of :15s was 100.9%. (The relative interest value compares the program presented with :15s against the program presented with :30s. It is calculated by dividing the respective measure for :15s by the measure for :30s.)

Analysis of variance also showed a significant interaction effect between commercial length and age ($F = 11.65$, df = 2,654, $p < .001$) and commercial length and education ($F = 18.14$, df = 2,654, $p < .001$).

Age. In terms of age, interest to watch future episodes as a function of :15s and :30s was highest among younger viewers, (Exhibit 11.16). However, the range between age groups was much greater as a function of :15s (3.44 − 2.66 = 0.78) than :30s (3.18 − 2.98 = 0.20). Relative interest (:15s/:30s) was about parity among the young (108%) and middle aged (106%), but was somewhat less among the older age group (89%).

Education. In terms of education (Exhibit 11.16), interest to watch future episodes as a function of :15s was similar for all viewers, ranging from a low of 2.90 (grammar-school-educated) to a high of 3.18 (college-educated). But as a function of :30s it was very dissimilar, ranging from a low of 2.08 (grammar-school-educated) to a high of 3.40 (high-school-educated). Relative interest (:15s/:30s) was well above parity among the grammar-school-educated (139%), about parity among the college-educated (106%) and was slightly lower among the high-school-educated (92%).

Hypothesis 2

Hypothesis 2 was not rejected. As hypothesized, curiosity to see the ending of the episode was significantly lower when the program was presented with :15s compared to :30s. Analysis of variance showed a significant main effect for commercial length ($F = 8.24$, df = 1,654, $p < .004$). Mean values

Exhibit 11.16
Interest, Curiosity, and Comparison

	Comm. Length* :15s	Comm. Length* :30s	Relative Value (:15s/:30s)
"INTEREST TO WATCH FUTURE EPISODES"			
Age**			
18-34	3.44	3.18	108%
34-49	3.32	3.14	106%
50+	2.66	2.98	89%
Education**			
College	3.18	3.00	106%
High School	3.13	3.40	92%
Grammar School	2.90	2.08	139%
"CURIOSITY TO SEE ENDING OF EPISODE"			
Age**			
18-34	2.75	2.30	120%
34-49	2.19	2.28	96%
50+	1.64	2.47	66%
Education**			
College	2.48	2.05	121%
High School	2.04	2.66	77%
Grammar School	1.97	2.04	97%
"COMPARISON TO SIMILAR PROGRAMS"			
Sex**			
Male	2.11	1.65	128%
Female	1.73	1.77	98%
Education**			
College	2.03	1.57	129%
High School	1.77	1.87	95%
Grammar School	1.77	1.54	115%

*Mean values

**Significant ANOVA interaction with commercial length ($p<.001$).

Note: "LIKELIHOOD TO RECOMMEND..." is not presented, because no significant ANOVA interaction occurred between commercial length and age ($p<.937$), sex ($p<.261$), or education ($p<.418$).

for curiosity to see the episode's ending were 2.36 for :30s viewers and 2.20 for :15s viewers. Among this national sample of typical adult television audience members, relative curiosity about the ending (:15s/:30s) was 93%.

Analysis of variance also showed a significant interaction effect between commercial length and age ($F = 30.36$, df = 1,654, $p < .001$) and commercial length and education ($F = 20.59$, df = 2,654, $p < .001$).

Age. In terms of age, curiosity to see the ending of the episode as a function of :15s was highest among younger viewers, while as a function of :30s it was highest among older viewers (Exhibit 11.16). Relative curiosity (:15s/:30s) was above parity among the young (120%), about parity among the middle aged (96%) and was substantially below parity among the old (66%).

Education. In terms of education, curiosity to see the ending of the episode as a function of :15s was highest among college-educated viewers, while as a function of :30s it was highest among high-school-educated viewers (Exhibit 11.16). Relative curiosity (:15s/:30s) was highest among the college-educated. In fact, relative comparison was above parity among the college-educated (121%), about parity among the grammar-school-educated (97%), and below parity among the high-school-educated (77%).

Hypothesis 3

Hypothesis 3 was rejected. Opposite than hypothesized, comparison of the program was significantly more favorable when the program was presented with :15s compared to :30s. Analysis of variance showed a significant main effect for commercial length ($F = 5.67$, df = 1,654, $p < .018$). Mean values for comparison of this program with similar others were 1.72 for :30s viewers and 1.87 for :15s viewers. Among this national sample of typical adult television audience members, relative comparison of the program (:15s/:30s) was 109%.

Analysis of variance also showed a significant interaction effect between commercial length and sex ($F = 13.67$, df = 1,654, $p < .001$) and commercial length and education ($F = 6.7$, df = 2,654, $p < .001$).

Sex. In terms of sex, comparison as a function of :15s was most favorable among males, while as a function of :30s it was most favorable among females (Exhibit 11.16). The relative comparison value (:15s/:30s) was well above parity among males (128%) and about parity among females (98%).

Education. In terms of education, comparison as a function of :15s was most favorable among college-educated viewers, while as a function of :30s it was most favorable among high-school-educated viewers (Exhibit 11.16). The relative comparison value (:15s/:30s) was highest among the college-educated. In fact, relative comparison was well above parity among the college-educated (129%), above parity among the grammar-school-educated (115%), and about parity among the high-school-educated (95%).

Hypothesis 4

Hypothesis 4 was rejected. Different than hypothesized, likelihood to recommend this program to friends was not significantly lower when the program was presented with :15s compared to :30s. Analysis of variance

did not show a significant main effect for commercial length ($F = .153$, df = 1,654, $p < .695$). Mean values for the likelihood to recommend this program to friends were 2.28 for :30s viewers and 2.31 for :15s viewers. Among this national sample of typical adult television audience members, relative likelihood to recommend this program to friends (:15s/:30s) was 101%. As Exhibit 11.16 reports, analysis of variance did not indicate any significant interaction effects for this variable.

DISCUSSION

This experiment achieved its goal. Realistic causal data were collected and the results offer insight into the impact that :15s have on editorial context. The conclusion is that the marketing discipline's fear about increasing use of :15s appears unjustified.

Presenting a program with :15s, rather than :30s, did cause viewers to perceive more commercials and more total commercial time. However, regardless of whether the program was viewed in a setting of :15s or :30s, the different formats did not change interest to watch future episodes or the likelihood to recommend the program to friends. In fact, program comparisons with similar programs was more favorable when presented with :15s than when presented with :30s.

An implication for marketing managers is that imbedding :15s in television programs does not necessarily adversely impact consumer behavior regarding the editorial context. Consumers are equally likely to tune in to future programs of a television program series, and to recommend that their friends do the same.

While other research reports that :15s are compared less favorably than :30s, this current research indicates that those comparisons do not generalize to the editorial context (which, if they did, could in turn indirectly lower effectiveness of subsequent commercials by being presented in less favorable states of mind). In contrast, the adverse impact on curiosity presents a negative implication for marketing managers. Since programs presented with :15s caused curiosity to see the program ending to be lower, these consumers are more likely to switch channels during a program and to even discontinue viewing during a particular program, which means that all the respective commercial breaks are less likely to be seen.

Behavior versus Cognition

Viewers exhibit similar "behavior responses" and dissimilar "cognitive responses." Behavior-related reactions to the program presented with :15s were equally favorable as when presented with :30s. Both groups of viewers showed equal interest to watch future program episodes (Hypothesis 1) and equal likelihood to recommend the program to friends (Hypothesis 4). Cognitive reactions to the program presented with :15s resulted in more favorable "competitive" comparisons than when presented with :30s (Hypothesis 3). Data regarding curiosity is the only adverse finding. Compared to the program presented with :30s, a program presented with :15s caused lower curiosity in the ending of an episode (Hypothesis 2).

In the marketplace, behavior responses and cognitive responses are interdependent. While a global favorable comparison may underlie consumer behavior to begin viewing a television program, a lack of curiosity about the program ending may underlie consumer behavior to interrupt and discontinue viewing the program before it is completed. Viewers may tune in a program on a regular basis, regardless of whether it is presented with :15s or :30s, but their interest during the program is more likely to wane with the former. Since curiosity to see the story ending is lower with :15s than :30s, these former viewers are more likely to switch channels during commercial breaks. Underlying feelings may be that "nothing important" will be missed if the story's end is not seen. Similarly, viewers may not stay tuned during the entire program.

State of Mind

Research shows generally negative cognitive reactions to :15s. These reactions regard quality (e.g., less persuasive) and quantity (e.g., clutter via more commercials and more perceived commercial time). It seems reasonable to assume that these negative reactions are associated with parallel state of mind. The less favorable state of mind may be expected to generalize from the interspersed commercials to their attending program. (Ultimately, this program-related state of mind could in turn be expected to negatively impact attending and subsequent commercial effectiveness.)

While intuitively appealing, the results of this experiment do not support this speculation. Reports that consumers describe :15s less positively than :30s appear limited to the commercials themselves. Certainly, the report here is that while consumers may make negative comparisons of :15s with :30s, these comparisons do not generalize to the attending program. In fact, a program presented with :15s is actually compared more favorably than when presented with :30s.

An explanation for this impact seems to be a consumer referencing effect. Consumers may be using commercials as a reference or a norm for their feelings. A less appealing commercial break (e.g., :15s), which the consumer may use as a norm for his or her state of mind, makes the attending contrasting program appear more appealing than it does when commercial breaks are more appealing (e.g., :30s).

An implication is that the circular "Catch-22" (commercial to program to commercial) discussed earlier does not appear to apply. At least not in the sequence of commercial to program. The evidence does not support

notions that a less favorable state of mind related to commercials generalizes to a less favorable state of mind related to the program (which ultimately might be expected to again generalize to a less favorable state of mind related to commercials). Rather, the evidence suggests that consumers discriminate rather efficaciously between state of mind during commercial breaks and state of mind during attending television programs.

Actions versus Thoughts

It is discomforting for marketing managers to buy commercial time on television programs that consumers compare less favorably to other programs. It appears initially that programs presented with :15s rather than :30s do not fit this situation. On closer scrutiny, marketers who buy time on programs presented with :15s appear to have less assurance of keeping the viewers for the entire program and for all the commercials. Viewers of these programs are likely to be less attentive, as indicated by the fact that :15s caused viewers in this experiment to be less curious about how the episode ends.

In contrast, marketing managers should be comforted when buying commercial time on programs with regular viewers. It appears that marketers who buy time on programs presented with :15s, versus :30s, have equal assurance for regular viewers from episode to episode. The reason may be that programs presented with :15s, versus :30s, are actually compared more favorably with other programs. Viewers of these programs are likely to be equally constant, as indicated by the fact that :15s did not cause viewers in this experiment to be less interested in watching future episodes of the program or in recommending the program to friends.

An implication regards marketing strategies that focus on, or at least include, repeated exposures (via constant viewers of a program). In these instances, price differences come into consideration. If a marketing strategy calls for repeated exposures (via constant viewers of a program), and if 15-second time slots cost less than 30-second time slots, it will be advantageous to allocate budgets accordingly.

SUMMARY/ABSTRACT

While 15-second commercials are reported generally to be about 80% as effective as 30-second commercials, substantial fear has been expressed about the negative impact of the 15-second commercials, via clutter, on overall advertising effectiveness. At the same time, no research has addressed the effect that 15-second commercials have on their respective editorial context; the attending television program itself.

Therefore, impact of :15s on editorial context was the focus of this experiment. The methodology involved an experiment with a national sample of 655 American adults. An actual television program, part of a network series not yet broadcast, was presented with commercial breaks composed of :15s or :30s. The data collected included: interest to watch future episodes, curiosity to see an episode's ending, comparison of the program with similar others, and recommendation of the program to friends.

In conclusion, this experiment suggests that fear of :15s leading to advertising clutter is unfounded. Results showed that :15s did not have a negative impact on "behavior responses" and had a mixed impact on "cognitive responses." The :15s, versus :30s, did not lower interest to watch future episodes of the program or lower the likelihood to recommend the program to friends. At the same time, :15s improved the comparison of the program with similar programs. The only negative impact of :15s was to lower curiosity to see the ending of an episode.

NOTES

1. Scott Ward, Terence A. Oliva, and David J. Reibstein, "Effectiveness of Brand-related 15-second Commercials," *Journal of Consumer Marketing* 11, no. 2 (1994), pp. 38–44.

2. Stephen K. Lonning, "The Fifteen-Second Commercial: A Viable Option," *Marketing & Media Decisions* (April 1988), pp. 100 and 102.

3. George S. Fabian, "15-Second Commercials: The Inevitable Evolution," *Journal of Marketing Research* 26 (August 1986), pp. RC3–RC5.

4. Gordon L. Patzer, "Multiple Dimensions of Performance for 30-Second and 15-Second Television Commercials," *Journal of Advertising Research* 31, no. 4 (August/September 1991), pp. 18–25.

5. Scott, Oliva, and Reibstein, "Effectiveness of Brand-related 15-second Commercials," pp. 38–44.

6. Surendra N. Singh, Michael L. Rothschild, and Gilbert A. Churchill, "Recognition versus Recall as Measures of Television Commercial Forgetting," *Journal of Marketing Research* XXV (February 1988), pp. 72–80.

7. Marvin S. Mord and Edith Gilson, "Shorter Units: Risk—Responsibility—Reward," *Journal of Advertising Research* 25 (August 1985), pp. 9–19.

8. "Emotional impact can cut clutter of 15-second spots," *Marketing News*, December 5, 1986, p. 13.

9. Ronald Alsop, "More Companies Squeeze Ads Into Bargain 15-Second Spots," *Wall Street Journal*, January 29, 1987, p. 29.

10. Betsy Sharkey, "Emotional Pitch Can Take the Chaos Out of 15s," *Adweek*, February 11, 1985, p. 31.

11. Verne Gay, "Split-30 Boom Expected: Is 15-Second Unit Next?" *Advertising Age* 55, no. 72 (October 25, 1984), pp. 2 and 54.

12. Mord Gilson, "Shorter Units: Risk—Responsibility—Reward," pp. 9–19.

13. Ibid.

14. William M. Claggett, "The Long View of The Short Commercial," *Journal of Advertising Research* 26 (August 1986), pp. RC11–RC12.

15. Mord and Gilson, "Shorter Units: Risk—Responsibility—Reward," pp. 9–19.

16. "Emotional impact can cut clutter of 15-second spots," p. 13.

17. Michael L. Ray and Peter H. Webb, "Three Prescriptions for Clutter," *Journal of Advertising Research* 26 (February 1986), p. 69.

Chapter 12

Analysis via Extensive Quantitative Tests and Techniques

Organizations concerned with survival must take a serious interest in the communication that flows *from* themselves to their audiences. Although the type of communication and type of delivery medium vary, a substantial amount of communication from both profit and nonprofit organizations is persuasive communication delivered via nonpersonal communication, of which a person's physical attractiveness is an integral factor. Therefore, the research problem of this current project was to design and perform an experiment to investigate the effects of communicator physical attractiveness upon the effectiveness of persuasive communication.

Two empirical questions were addressed. First, are communicators of different levels of physical attractiveness perceived differently in regard to source credibility? Second, what is the relationship between communicator physical attractiveness and persuasive communication effectiveness, as measured in terms of receivers' recall, attitudes, beliefs, perceptions, and evaluations. Because the focus of this chapter is on experiments as a research methodology in marketing, the vast amount of related research regarding the psychological, sociological, and marketing aspects of a person's physical attractiveness, as well as the communications research, is excluded from this current discussion.[1]

In response to the above research problem, the objectives of this experiment were to perform research that (1) corrects procedural problems characteristic of past research and (2) addresses, through appropriate hypotheses, the empirical questions stated in the above paragraph. Also, all hypotheses were theoretically based, and where possible, substantiated by existing empirical knowledge.

OVERVIEW

A $2 \times 2 \times 4$ factorial experiment design was used for this research project. The three manipulated independent variables were: (1) the sex of the receiver (male and female), (2) the sex of the communicator (male and female), and (3) the physical attractiveness of the communicator (no-photo, low, moderate, and high). To test the seven research hypotheses, male and female receivers (subjects) were shown a printed advertisement mock-up with either a male or female communicator representing one of the four physical attractiveness conditions.

The identical advertisement mock-ups were presented in each condition, with only the photograph element manipulated. These photograph manipulations involved: (1) a photograph of either a male or female communicator of either low, moderate, or high physical attractiveness and (2) no photograph at all. The low, moderate, and high physical attractiveness levels were operationalized from the results of an earlier rating process consisting of a truth-of-consensus judging method, which is detailed in the following "Phase One" section of this chapter. The dependent variables were the subjects' responses in a questionnaire assessing their recall, perceptions, evaluations, attitudes, and beliefs. Persuasive communication effectiveness was operationalized to be response differences exhibited in regard to the experimental treatment differences.

This research was performed in two major phases. The primary purpose of phase one was to obtain photographs of individuals who represented low, moderate, and high levels of physical attractiveness for each sex. The purpose of phase two was to investigate the relationship between communicator physical attractiveness and persuasive communication effectiveness. Based on the operational definitions of communicator physical attractiveness and persuasive communication effectiveness, this investigation used the photographs obtained in phase one to test the seven hypotheses.

Phase One

Physical attractiveness was defined as the degree to which a stimulus person is pleasing to observe. Based on this definition, subjects were presented with photographs of individuals and asked to judge their physical attractiveness. This judging process involved two different methods, with each being administered at two different times (i.e., test-retest). The two methods were: (1) a 7-point bipolar rating scale and (2) an assimilation-contrast grouping task. To determine the validity of the physical attractiveness construct measures, reliability coefficients were calculated on the test-retest measures for each method.

The reliability coefficients proved sufficiently large to permit (and to justify) using the collected data as a basis to select stimulus persons to later serve as communicators. To accomplish this selection of communicators to represent the three levels of physical attractiveness, the mean score and standard deviation for each stimulus person were calculated. Then, using t-tests with these data, three stimulus persons each for the low and high communicator physical attractiveness levels and two stimulus persons for the moderate communicator physical attractiveness level were selected for each sex.

Phase Two

Phase two began with a pilot test to determine: (1) if procedural problems existed and (2) if the physical attractiveness levels identified in phase one would be maintained when combined with the advertisement mock-up. The successful pilot test results permitted (and justified) beginning the actual experiment. The methodology used in this experiment involved numerous procedures to assure internal validity, that is, efforts were made (and justified) to maximize the likelihood that the results, whatever they would be, were caused by the experimental manipulation of communicator physical attractiveness, and not by some other variable.

PHASE ONE

Multitrait–Multimethod Matrix

To operationally define the physical attractiveness of communicators to be used in phase two, a modified form of the multitrait–multimethod matrix was constructed. The two traits measured, for exploratory purposes, were sexiness and physical attractiveness. The two methods used were a bipolar rating scale method and an assimilation-contrast grouping method. These methods were replicated two weeks apart, and no subjects who served in the rating method also served in the grouping method, and vice versa.

The purpose of the multitrait–multimethod matrix (MTMM) was to strengthen the physical attractiveness construct. The MTMM matrix procedure is not new, but was first proposed in 1959 by Campbell and Fiske in the psychological literature,[2] and more recently it was advocated by others in the marketing literature.[3] However, the validity and reliability of measures assessing the physical attractiveness construct within persuasive communication have not been addressed in any of the existing research, and have been given only limited attention in the general physical attractiveness research.

Definition of Traits

The research literature gives little attention to a definition of sexiness and no attention to a definition of physical attractiveness; however, definitions for these terms were used in this experiment. A modified version of Freud's definition of sexiness and an original definition of physical attractiveness were used:

Sexiness. The extent to which the appearance of a stimulus person arouses a sexual or an erotic idea in the observer's mind. With opposite-sex dyads this arousal may be directly related to oneself, and with same-sex dyads this arousal may be the observer's perceptions of others' reaction.

Physical Attractiveness. The degree to which a stimulus person is pleasing to observe.[4]

Measurement Methods

There were substantial differences between the two methods used. First, the bipolar rating scale method required the subjects to view and rate only one stimulus person at a time, while the assimilation-grouping method required the subjects to view and group all the stimulus persons at the same time. Second, the instructions and forms for each method differed accordingly. Third, different subjects were used for each method.

The bipolar rating scale was a 7-point, labelled, continuum ranging from extremely low to extremely high physical attractiveness or sexiness as appropriate. Sixty subjects (30 males and 30 females) used this bipolar rating scale to rate the physical attractiveness of both male and female stimulus persons. Once the rater completed the physical attractiveness ratings, he/she was given a rest for approximately 10 minutes, after which the rater was given a different sequence of the same photographs and asked to rate them for each stimulus person's sexiness. For each rating the photographs were randomized separately for each subject. In addition, the first sex to be rated was also randomly assigned to each subject. To capture the essence of both physical attractiveness and sexiness, the definitions of each concept were printed on the instruction sheets, the response forms, and were in front of the subject at all appropriate times.

The assimilation-contrast grouping followed a "7-level distribution," and used 60 subjects (30 males and 30 females) who were not used in the bipolar rating scale method. The judges here were given all the photographs and asked to rank the stimulus persons by forming groups through comparison and contrast. To perform their task the subjects were provided a labelled form with equal intervals. The photographs of stimulus persons placed within a group were to represent the same or similar levels of physical attractiveness and sexiness, while the photographs placed in different groups were to represent different levels of physical attractiveness and sexiness, respectively. The judges were asked to form different groups in columns that ranged from extremely low to extremely high physical attractiveness and sexiness, respectively; however, a forced distribution was not used (i.e., subjects were permitted to use as many or as few groups as they felt appropriate). Once the judge finished the physical attractiveness grouping, he/she was given a rest for approximately 10 minutes. Then the judge was given a different sequence of the same photographs and asked to group them for each stimulus person's sexiness. For each grouping the photographs were randomized separately for each subject. In addition, the first sex to be judged was also randomly assigned for each subject. To capture the essence of physical attractiveness and sexiness with the grouping process, the definitions of each concept were printed on the instruction sheets and were in front of the subjects at all appropriate times.

Test-Retest

Measures of reliability for the multitrait–multimethod matrix were obtained through a test-retest, two weeks apart. This method and time period are suggested, as a generally recommended procedure, by Peter (1979, p. 8). Churchill (1979, p. 70) feels that a test-retest should not be used because the subjects' memory will cause subjects to reply in the same way the second time as the first. For this research, Churchill's concern is less relevant because of the large number of photos (42) for each judgment process and the random order for each administration. Because of this design, it is unlikely that the subjects could recall their responses from two weeks earlier.

Multitrait–Multimethod Matrix Results

The results of the MTMM procedure are presented in modified MTMM matrix form in Exhibit 12.1 for the male subjects and in Exhibit 12.2 for the female subjects. The Cronbach's Alpha Reliability Coefficients obtained in the current experiment (.61 to .82) appear very good and indicate the measures of the physical attractiveness construct were quite reliable for both methods.[5] This reliability was exceptionally high for the rating method and somewhat lower for the grouping method. The reliability coefficient value was smallest (.61) for the male subjects using the grouping method to judge female photographs. The validity results are not presented in these MTMM matrices because of the high interdependence and similarity of the constructs used, and because of a lack of clarity as to whether the subjects responded to the sexiness ratings as they perceived or as they perceived others perceive. The consequence of this action is that the validity com-

Exhibit 12.1
Reliability Coefficients for the Physical Attractiveness Construct: Based on Responses from the Male Judges

	METHOD ONE (RATING) TIME ONE	METHOD TWO (GROUPING) TIME ONE
METHOD ONE (RATING)	.791 (MP .744 (FP	
METHOD TWO (GROUPING)		.773 (MP .609 (FP

NOTE:

(1) "MP" represents male photographs
"FP" represents female photographs

(2) Coefficient values are Cronbach's Alpha Reliability Coefficient calculated with the SPSS Computer Package.

ponent of the MTMM matrix regarding the physical attractiveness construct was not investigated in this experiment.

Subjects

A total of 120 subjects (60 males and 60 females) made the required judgments in phase one. Thirty males and thirty females were used for the bipolar rating scale while thirty different males and thirty different females were used for the assimilation-contrast grouping. All subjects were run under the same setting and the same conditions. These conditions were similar rooms with similar furnishings, lighting, stimulus materials, verbal and written instructions, assurance of complete anonymity in regard to their responses, and the same experimenter.

Stimulus Materials

The stimulus materials were black-and-white photographs measuring 1½ inches by 2 inches. Each photograph was of a college senior in his/her early twenties, taken from a university yearbook. Only the faces were pictured and no stimulus person was selected who wore glasses, a beard, a mustache, sideburns longer than mid-ear, or had a blemished complexion. All individ-

Exhibit 12.2
Reliability Coefficients for the Physical Attractiveness Construct: Based on Responses from the Female Judges

	METHOD ONE (RATING) TIME ONE	METHOD TWO (GROUPING) TIME ONE
METHOD ONE (RATING) TIME TWO	.821 (MP .776 (FP	
METHOD TWO (GROUPING) TIME TWO		.688 (MP .686 (MF

NOTE:

(1) "MP" represents male photographs
"FP" represents female photographs

(2) Coefficient values are Cronbach's Alpha Reliability Coefficient calculated with the SPSS Computer Package.

uals were pictured in the same position, size, and location for all subjects (both in phase one and phase two). To control for potential effects due to clothes, all photographs were from the upper neck region and up.

College Yearbook. The photographs were chosen from a college yearbook because these were the most standardized stimulus materials among the alternatives (e.g., personal snapshots or professional photographs from advertising and talent agencies). With the yearbook photographs the stimulus persons were all pictured in similar dress and grooming (because photographs showed only the individual's upper neck area and face), similar position, and similar facial expression. In addition, the photographs were the same size (except for minor variation that was due to adjusting the photograph to ensure the only body area visible was the face), used the same background, had the same photographer using the same equipment, and could be controlled for appearance (e.g., hair length, facial hair, complexion, and eyeglasses). This extreme standardization allowed strict control over variables that otherwise might influence ratings of physical attractiveness. The result of this control permitted the research outcome to be more readily attributed to the stimulus person's physical attractiveness rather than to some other variables (e.g., pose, hair, photo background, or photographer's skill and style). To place such high emphasis on standardized photographs is consistent with Weber and Cook (1972, p. 293), who contend that a major goal of experimental designs should be to minimize differences in observable characteristics. In other words, the best experimental treatments should differ only in a single value, that being the independent variable manipulation.

In addition to controlling for appearance factors, the effects of subjects knowing the stimulus persons were also controlled. This control was performed in three ways. First, to minimize potential recognition and/or acquaintance factors, stimulus persons were selected from a college yearbook representing a university population of 20,000 students. Second, four academic years separated the subjects and stimulus persons; consequently, the stimulus persons usually had graduated before the subjects even began college. Third, both written and oral instructions asked the subjects to indicate, on their rating forms, if they recognized or knew any of the stimulus persons. Data from subjects indicating any sort of recognition or acquaintance were omitted from analyses. This inquiry procedure identified only one subject who thought he recognized one of the female stimulus persons. Based on these results, it was concluded that recognition and acquaintance factors were controlled successfully.

Faces and Stimulus Materials. Studying faces or using faces as stimulus materials in marketing communications research is appropriate because of the properties a face possesses. The discussion here first compares face stimuli with nonface stimuli, and then compares the face stimuli with other faces. First, faces are more recognizable than other patterns, such as ink blots, photographs of show crystals, houses, stick figures, and other designs. Second, the characteristics displayed in photographed faces serve as a better aid to recall than the processes of controlled imagery. Third, visual information serves as a greater determinant of impression formation than does verbal information.

Because different faces possess different properties, perceptions of specific faces may differ accordingly. In a task recognition study, photographs of individual's faces judged earlier to be either low or high in physical attractiveness were recognized better than the individuals (faces) judged earlier as moderate in physical attractiveness. In another study involving faces and memory, faces were found to differ in memorability as a function of their physical attractiveness—those faces earlier judged as physically attractive were recognized with greater frequency than those faces not judged so.

Procedure. The photographs of 42 college seniors, as they appeared in their college yearbook, were reproduced; consequently, all persons were similarly posed, dressed, and groomed. A conscious attempt was made to choose an approximately equal number of people representing low, moderate, and high levels of physical attractiveness for each sex. For the subjects who served as judges, the photographs were randomized for each task for each judge. This randomization was to counterbalance against possible boredom and/or practice effects; a table of random numbers was used.

When the subject (judge) arrived, he/she was seated and given a letter of introduction. Following this letter was an instruction page telling the judge what to do and asking him/her to rate two sample photographs (one male and one female). After completion of the practice exercise, the instructions asked the judge to indicate to the experimenter if he/she understood what to do. At all times the amount of verbal instruction was kept to a minimum, and involved only instruction about the rating mechanics (i.e., a serious attempt was made to eliminate all potential bias and/or demand characteristics).

Phase One Results

To select photographs for the advertising mock-ups, the data from the rating and grouping methods were used. The scales used for these two methods were assigned numerical values ranging from one to seven and were treated as interval data. On the basis of the physical attractiveness scores, three photographs of each sex were selected to represent the high and low levels of physical attractiveness, and two photographs of each sex were selected to represent the moderate physical attractiveness level.

The intent was to select more than one photograph per physical attractiveness level. The specific number was not determined until after the statistics were calculated and the likely photographs emerged. The reason multiple stimulus persons were chosen to represent each physical attractiveness level was to minimize any unique characteristics or unique effects that a specific stimulus person may have possessed. In the experimental treatment procedures, each photograph, representing a specific physical attractiveness level, was presented an equal number of times. Despite the potentially misleading impact that a unique stimulus person could have on the results, especially at the extremes of the physical attractiveness continuum, only one study in the physical attractiveness and persuasive communication research has incorporated multiple stimulus persons into the research design.[6] A much smaller percentage appears to have done so in the general physical attractiveness research.

Once the judges' data were collected, the mean score and standard deviation for each stimulus person were calculated for each of the two methods and for each of the four judging periods. Next, the means representing the lowest, middle, and highest scores were grouped together for each sex. From each of these groups those photographs with the smallest standard deviation were selected to represent their respective level of physical attractiveness. Finally, t-tests were performed to ensure that significant differences existed between the scores for photographs that were in different physical attractiveness levels. These mean scores and standard deviation

values are presented in Exhibit 12.3 for the photographs of male stimulus persons, and Exhibit 12.4 for the photographs of female stimulus persons who were selected to serve as communicators in the experimental treatment conditions of phase two.

PHASE TWO

Pilot Test

When phase one was completed, the research project moved into a pilot test of the actual experiment. Thirty additional subjects were presented with the instructions, advertisement mock-ups, questionnaires, and in total, the entire procedure for phase two. These "dress rehearsal" data were not analyzed in the final research write-up, but were used to attempt to identify potential weaknesses and problems in methodology before the experiment was actually conducted. This objective was accomplished by constantly monitoring and examining the data collections, as well as carrying out both formal and informal posttest discussions with the pilot test subjects. Because no problems were identified, and no changes appeared necessary, the actual experiment was begun.

Subjects

The data from a total of 360 subjects (180 males and 180 females), who had not served in any part of the experiment thus far, were used as the sample for phase two. All subjects were run under the same setting and the same conditions. (In addition, phase two was run as similarly to phase one as was technically possible.) The subjects were college juniors and seniors from introductory business courses.

College Students as Subjects. Although nonstudents may be desirable subjects, college students were certainly of considerable value. This value was established through a number of explanations. First, most research involving physical attractiveness has used college populations. This popular practice may be viewed as "too narrow"; however, it can also be viewed as quite good because it lends itself to the advantage of comparability and continuity.

Second, discussions of external validity questions in business and other social science research often conclude that, dependent on the situation, college students are permissible subjects. These affirmative discussions can be summarized by a quote from W. Oaks:

> (R)esearch with college students as subjects is just as valid as research drawing on any other subject population. A behavioral phenomenon reliably exhibited is a genuine phenomenon, no matter what population is sampled in the research in

Exhibit 12.3
Statistics of Male Stimulus Persons Selected to Be Communicators for Experimental Treatments

	RATING METHOD N=60	GROUPING METHOD N=60	AVERAGE OF RATING AND GROUPING METHODS N=120
LOW PHYSICAL ATTRACTIVENESS			
	M = 1.38 SD = 0.64	M = 1.18 SD = 0.42	M = 1.28 SD = 0.53
	M = 1.97 SD = 0.88	M = 1.72 SD = 0.92	M = 1.85 SD = 0.90
	M = 2.03 SD = 0.79	M = 1.77 SD = 0.91	M = 1.90 SD = 0.85
MODERATE PHYSICAL ATTRACTIVENESS			
	M = 3.93 SD = 1.04	M = 3.92 SD = 1.25	M = 3.93 SD = 1.15
	M = 3.81 SD = 1.10	M = 4.33 SD = 1.13	M = 4.07 SD = 1.12
HIGH PHYSICAL ATTRACTIVENESS			
	M = 4.63 SD = 0.85	M = 5.08 SD = 1.06	M = 4.86 SD = 0.96
	M = 4.87 SD = 1.06	M = 5.17 SD = 1.43	M = 5.02 SD = 1.25
	M = 5.89 SD = 0.86	M = 6.26 SD = 0.84	M = 6.08 SD = 0.85

NOTE:

(1) If no significant differences for photos existed between the test and retest scores of either methods, statistics in column 3 were used to identify the physical attractiveness level of the stimulus person.

(2) Based on column 3 statistics, all photos in different levels were significantly different at .000 level of probability.

which it is demonstrated. . . . No matter what population a researcher samples, whether it be psychology students, real-people volunteers, public school students, or whatever, there are probably some behavioral phenomena that would be manifested differently in that population due to an interaction effect of the particular characteristic of that subject population.

This suggests, then, that the generalizability of the results of behavioral research is not a function of the population sampled, but rather that the external validity of the research depends on the interaction of subject characteristics and the particular behavioral phenomenon with which one is concerned. (pp. 961–962)[7]

Exhibit 12.4
Statistics of Female Stimulus Persons Selected to Be Communicators for Experimental Treatments

	RATING METHOD N=60	GROUPING METHOD N=60	AVERAGE OF RATING AND GROUPING METHODS N=120
LOW PHYSICAL ATTRACTIVENESS			
	M = 1.27 SD = 0.56	M = 1.12 SD = 0.35	M = 1.20 SD = 0.46
	M = 2.09 SD = 0.94	M = 1.73 SD = 0.79	M = 1.91 SD = 0.87
	M = 2.15 SD = 1.00	M = 1.58 SD = 0.81	M = 1.87 SD = 0.91
MODERATE PHYSICAL ATTRACTIVENESS			
	M = 3.85 SD = 0.99	M = 5.70 SD = 1.00	M = 5.52 SD = 1.00
	M = 3.99 SD = 1.18	M = 4.08 SD = 1.47	M = 4.04 SD = 1.33
HIGH PHYSICAL ATTRACTIVENESS			
	M = 5.33 SD = 0.99	M = 5.70 SD = 1.00	M = 5.52 SD = 1.00
	M = 5.43 SD = 1.04	M = 5.68 SD = 1.17	M = 5.56 SD = 1.11
	M = 5.49 SD = 0.89	M = 5.86 SD = 1.05	M = 5.68 SD = 0.97

NOTE: (1) If no significant differences for photos existed between the test and retest scores of either methods, statistics in column 3 were then used to identify the physical attractiveness level of the stimulus person.

(2) Based on the statistics in column 3, all photos in different levels were significantly different at the .000 level of probability.

Third, there is the issue of to what population the results of this research were intended to apply. Disregarding study limitations for a moment, the current experiment was intended to generalize to individuals in the "consumer market." Consequently, college students were appropriate in several respects: (a) college students may be inappropriate to generalize to a distinctly different behavior or a different population, such as business decisions of industrial executives, yet, they may well be appropriate for similar behavior by similar populations; (b) this research is not "the final chapter," rather, it is a "first step" to be later followed by research with different populations to, in fact, prove or disprove generalizability; (c) the product advertised (headache and minor pain reliever) is a product bought and used by the subjects, and is a product shown not to elicit different familiarity and usage responses between students and nonstudents; and (d) a theoretical phenomenon is under study, and the relevant theory does not differentiate between populations.

Stimulus Materials

The stimulus materials were black-and-white advertisement mock-ups printed on 8½ × 11-inch paper. The advertisement mock-ups consisted of a male or female communicator of either low, moderate, or high physical attractiveness. The advertisement mock-up was for a new, nonexistent headache and minor pain reliever. As a control for physical attractiveness, a fourth treatment consisted of no physical attractiveness information (i.e., no photograph), combined with the advertising message.

In addition, baseline measures were obtained by controlling for all treatment effects. This control was achieved by requesting the subjects in this experimental condition to complete a modified form of the experimental questionnaire, but without being presented with any treatment. Because these baseline measures were for exploratory purposes, to be used in additional analyses later, these data were not analyzed for this current experiment.

Printed Stimulus Materials. Hand-held photographs rather than projected slides, and printed materials rather than verbal messages were selected to allow for maximum individual receiver differences. Presenting slides in a group setting implies (and assumes) that subjects are equally interested and equally efficient in receiving the advertisement information. Such slide procedures do not allow for individual differences, while individually administered photographs and printed materials do. The individual subject in the latter method can view the advertisement mock-up, read the copy, and in general process the information as long as desired.

Printed stimulus materials individually administered and slide-show presentations administered to groups appear to address different issues that must be considered. If the communicator's physical attractiveness affects interest, the printed material procedure permits such an effect to operate through longer information processing time, which may result in different responses to the dependent measures. However, this effect is not permitted to operate in a group slide-show because the interested subject and the less interested subject are both exposed for the same amount of time. On the other hand, communicator physical attractiveness may affect the speed or ability of the subjects, such that information processing is done much faster and/or much more effectively, dependent on communicator physical at-

tractiveness. If this situation exists, the printed material procedure would not be as appropriate as the experimenter-controlled group slide-show. In summary, the current experiment addressed the first concern by allowing individual information processing time to vary, while future research efforts may address the second. Regardless of the concern addressed, this procedural decision must be attended to and cannot just be assumed the same, for it is likely that each of the two procedures tap different information processing aspects.

Procedure

There were 14 sessions of approximately 30–40 subjects in each session. To reduce potential experimental bias and/or demand characteristics, all the different experimental treatments were administered within each session. Individual subjects were randomly assigned to an experimental treatment, controlling for equal cell sizes. A total of 18 experimental conditions or cells were used (i.e., 2 communicator sexes by 2 receiver sexes by 4 levels of communicator physical attractiveness, plus 2 cells created by the control).

All subjects were given as identical research materials as possible. These materials consisted basically of: (1) an introduction letter, (2) an advertisement mock-up page, and (3) a response booklet made up of questions concerning the advertising, the product, the communicator, and the receiver (subject) as well as a postexperimental, open-ended, question to identify possible subject bias. One of the questions about the communicator served as a check on the manipulation of the physical attractiveness variable. The only variation of materials for treatment groups was with the communicator sex and communicator physical attractiveness. The difference between the treatment and control groups was minimized at all times.

When the subject arrived he/she was seated and given a file folder containing the appropriate research materials. The only verbal instructions given to the subjects were as follows:

> First, please don't open your folders until you are told to do so. Second, different people in this room will have different tasks that involve more time or less time than others; so people will finish at different times. Are there any questions? OK, you may now begin by opening your folder, reading the first page, and then continuing on as the instructions state.

The above words in quotation marks were written on the blackboard in each room, and pointed to as they were mentioned by the experimenter.

Manipulation Check. A manipulation check was included to determine if the experimental manipulations of communicator physical attractiveness were successful. It is poor research to assume a priori operationalizations and/or successful pilot studies of the independent variable manipulations will have the intended impact in the actual experiment. The manipulation check was placed toward the end of the questionnaire. This placement decision was based on the notion that the strongest causal inferences can be made when the dependent variable measures are taken before the manipulation check. Placement of this manipulation check in an earlier part of the questionnaire could have caused an unwanted subject awareness or suspicion about the experiment.

Subject Apprehensiveness. The research literature identifies a number of subject roles, as well as means to control these roles. The procedure of this experiment was designed to be consistent with recommendations purporting to minimize the negative effects of subject roles, with the most widespread role probably being apprehensiveness.

Although the apprehensive role is a particularly difficult behavior with which to deal, several actions exist to reduce apprehensiveness and increase valid inference. These actions, which were built into the current methodology, include: (1) low emphasis of experimenter, (2) nonexperimenter control over rewards and punishments, and (3) obvious subject response anonymity. Also important is that the experiment be not viewed as unimportant by the subjects.

In addition to satisfying the specified recommendations, the instructions were printed and read by the subjects, and treatments were delivered simultaneously, in an effort to minimize experimental bias, hypothesis learning, and demand characteristics. To meet potential suggestions of low experimenter status while avoiding the appearance that the research was trivial, the subjects were informed that their participation was needed for part of a doctoral dissertation involving a year-long research project. Furthermore, the introduction letter was typed on stationery with the university letterhead for the "College of Business, Graduate Programs in Business Administration." The control of the subjects' rewards and punishments was not in the hands of the experimenter; as a result, the nonexperimenter control recommendation was met. This lack of control was achieved by employing subjects (students) who were not related to the experimenter (i.e., students from outside the experimenter's classes). In addition, although this lack of control was not explicitly stated to the subjects, it appeared to be further achieved by the (1) absence of a consequent that rewarded or punished a subject's performance, and (2) anonymity of individual subjects. Finally, anonymity was accomplished by stressing to the subjects that: (1) individual responses would not be reported, but only aggregates, (2) they were not put to any time of identification on their response booklets, and (3) they were to place their completed response booklets into a sealed box with all the other subjects' booklets.

Faithful Subject. A faithful subject role is not necessarily bad for an experiment. While a faithful subject may reduce the external validity, the primary goals of an experiment are to make valid causal inferences and ensure that these inferences are not affected by a faithful subject. Although there appears to be no problem with a faithful subject role in an experiment, it is dealt with in the following section on General Artifact Control.

General Artifact Control. This section discusses the general experimental procedures that were used to minimize demand characteristics, subject bias, and general artifacts. A major source of artifacts in a laboratory experiment is the fact that subjects know they are in an experiment. But, if the subjects' attention can be diverted from the demand characteristics to the naturalistic social demands of the exercise, artifacts will not be significant.

The general procedure that minimized demand characteristics in the current experiment was actually the central topics or elements of the research. The author had previously observed, from subjects' comments during and after past research, that subjects become very interested and greatly involved with both advertising mock-ups and the physical attractiveness of a stimulus person. The result of this substantial interest and involvement is subjects who become absorbed in their activity, and oblivious to role behavior.

Subject Attitudes. Several questions were asked about the subjects' academic major and attitudes toward business research, advertising, and the product advertised. These questions were exploratory and may be used in additional research to identify relationships between the subjects' responses to those questions and responses to specific questionnaire items. This decision to include these questions also served as a precautionary measure, in the event that the experimental manipulations failed and it became necessary to determine if certain attitudes or academic majors tended to contaminate the data. The experimental treatments were successful in this study, so it was not necessary to use these additional measures.

Postexperimental Query. In the event that unexpected artifacts slipped into the research, a postexperimental query was used to identify such problems. The query was used to identify subjects who had learned the hypotheses, adopted a subject role, or created other subject bias. Because the postexperimental query revealed nine subjects who suggested awareness of the experiment's communicator physical attractiveness manipulations, these data were omitted from statistical analysis. A secondary benefit was that if no artifacts had been discovered, the procedure could be used to refute speculation of experimental bias in this area.

Design

An experimental $2 \times 2 \times 4$ factorial design was used. The three independent variables manipulated were: (1) the sex of the receiver, (2) the sex of the communicator, and (3) the physical attractiveness level of the communicator. The need for the experimental design is often stressed in marketing research. In the past, marketing has relied primarily on nonexperimental research, but to identify causal relationships, it appears that marketing research must use the experimental research design. An example of this need is a quote (Baker and Churchill) that emphasizes the limitation of past marketing communications research, which is a major reason why this current study used an experimental design:

> The severest problem in sales-effect research has been that most of the studies are nonexperimental, i.e., based on econometric techniques using cross-sectional or longitudinal data. Yet, it is well known that sales depend on a wide range of factors such as the actions of competitors, conditions in the industry or in the general economy, government regulations, the weather, or the firm's own distribution, selling, and pricing policies; none of which are controlled in the nonexperimental designs often used to measure sales effects.[8] (p. 209)

Both general and specific arguments can be given to support an experimental design in this current physical attractiveness and persuasive communication research. In general, the study of variables through experimentation allows greater depth, detail, and precision of measurement than is possible through other means of study. Four reasons for an experimental design for research specifically regarding the relationship between physical attractiveness and persuasive communication research are: (1) past research demonstrates this procedure is a suitable method to manipulate the physical attractiveness variable without affecting either internal or external validity, (2) past research shows that physical attractiveness possesses a causal relationship with other variables, (3) the experimental design allows for analysis of interaction effects among independent variables, and (4) people are not aware or are not willing to acknowledge the influence that physical attractiveness has on them, and the experimental procedure permits the disguising of the manipulation of physical attractiveness.

Data Analysis

To test the hypotheses, the data collected from phase two procedures were analyzed using a $2 \times 2 \times 4$ factorial analysis of variance, and Pearson Product-Moment Correlations. Multiple comparison techniques (Tukey's W Procedure and Duncan's New Multiple Range Test) were used to further analyze all significant results indicated by the analysis of variance.

Before performing the analysis of variance tests on the attitude components (for Hypothesis 3) and on the belief component (for Hypothesis 4), the data were subjected to a number of preliminary statistical tests. These preliminary analyses were done to test the validity of each component by determining if (1) the items used for each component were homogeneous, (2) the items within a component related to the intended component, and

(3) the items intended for one component were not strongly related to another component not so intended. These preliminary tests included: (1) factor analysis, (2) reliability coefficient alpha, (3) item-to-total correlations, (4) multivariate analysis of variance, and (5) interitem correlation coefficients. Once these preliminary tests were completed, analysis of variance was performed: (1) on the manipulation check, (2) separately on each of the three attitude components, (3) on the belief component, (4) on the perceived trustworthiness, expertise, and liking for the communicator, and (5) on the recall scores. Pearson Product-Moment Correlation was performed between perceived physical attractiveness and perceived product quality, product price, and product uniqueness.

RESULTS AND ANALYSIS

Because the analyses varied, depending on the problem addressed, the remaining portion of this chapter is organized into preliminary procedures, manipulation check results, and procedural discussions and results for each hypothesis as appropriate. While the results were not strictly as hypothesized, due often to interaction effects, a synoptic statement would have to be that, in general, the results were very supportive of the hypothesized effects of communicator physical attractiveness.

Preliminary Procedures

After the data were collected the questionnaire forms were scrutinized to identify potential bias. This screening was intended to eliminate questionnaires which contained either (1) items not answered or (2) responses to the postexperimental query that indicated a subject took part in an earlier phase of the research and/or knew the purpose of the research. In addition to the above questionnaire eliminations, questionnaires were randomly deleted to achieve equal cell sizes of 20 subjects per experimental condition. Due to either incomplete forms or subject knowledge of the research purpose, 31 of the 432 questionnaires collected were eliminated. Subject knowledge, when suggested, appeared not to be a function of the research design but rather a function of the subject's exposure to an earlier phase of the research or to word-of-mouth communication about the research.

The initial data collection of 24 subjects for each of the 18 experimental conditions produced a total of 432 questionnaires. Dependent on the number omitted due to incomplete forms and/or subject bias, a maximum of 4 questionnaires were randomly omitted from each cell. Omitting these 72 questionnaires (31 due to the problems just mentioned and 41 due to random omission for equal cell sizes) resulted in 360 remaining questionnaires. Finally, because the exploratory data were not used in analyses from the two cells which involved neither an experimental treatment, the communicator physical attractiveness variable, nor the communicator sex variable, another 40 questionnaires were deleted from the remaining 360. The end result of these subtractions was 320 questionnaires that represented 20 subjects in each of the 16 experimental conditions of the 2 × 2 × 4 factorial design.

Check of the Physical Attractiveness Manipulation

The question designed as a manipulation check of the physical attractiveness variable (the seventh question on the third page of the subjects' questionnaire booklet) asked the subjects to evaluate the physical attractiveness of the spokesperson for the product's advertising. Because of the physical attractiveness control conditions that did not use a photograph of a communicator, these data were necessarily omitted from the manipulation check analysis. Omitting those subjects who were not exposed to a photograph of a communicator involved four control conditions that reduced the number of experimental cells from 16 to 12, and the total number of questionnaires from 320 to 240 for the manipulation check only.

Exhibit 12.5 presents the results of the analysis of variance for the physical attractiveness manipulation check. Because there were no interaction effects, only the one significant main effect revealed in the ANOVA table is illustrated in Figure 12.1. The cell means presented in this figure were all significantly different from each other at the .05 level for both the Duncan and Tukey multiple comparison tests. Based on the ANOVA summary table that indicated no interaction effects and the consequent multiple comparison tests, the physical attractiveness manipulation appears to have been successful.

Hypothesis 1: Communicators of higher levels of physical attractiveness will be perceived as more trustworthy and of higher expertise than communicators of lower levels of physical attractiveness.

The results support this first hypothesis: Communicators of higher levels of physical attractiveness were perceived as more trustworthy and of higher expertise than communicators of lower levels of physical attractiveness. By asking the subjects to evaluate the trustworthiness and expertise of the spokesperson for the product's advertising, item numbers 5 and 6 on the third page of the subjects' questionnaire booklet were designed to assess perceived trustworthiness and expertise of the communicator used in the different experimental conditions.

The analysis of variance results for trustworthiness and expertise are presented in Exhibits 12.6 and 12.7, followed by the presentation of the mean scores in Figures 12.2 and 12.3, respectively. Both Exhibit 12.6 and Exhibit 12.7 show that the source of variation was due to only main effects

Exhibit 12.5
Analysis of Variance of the Physical Attractiveness Manipulation Check

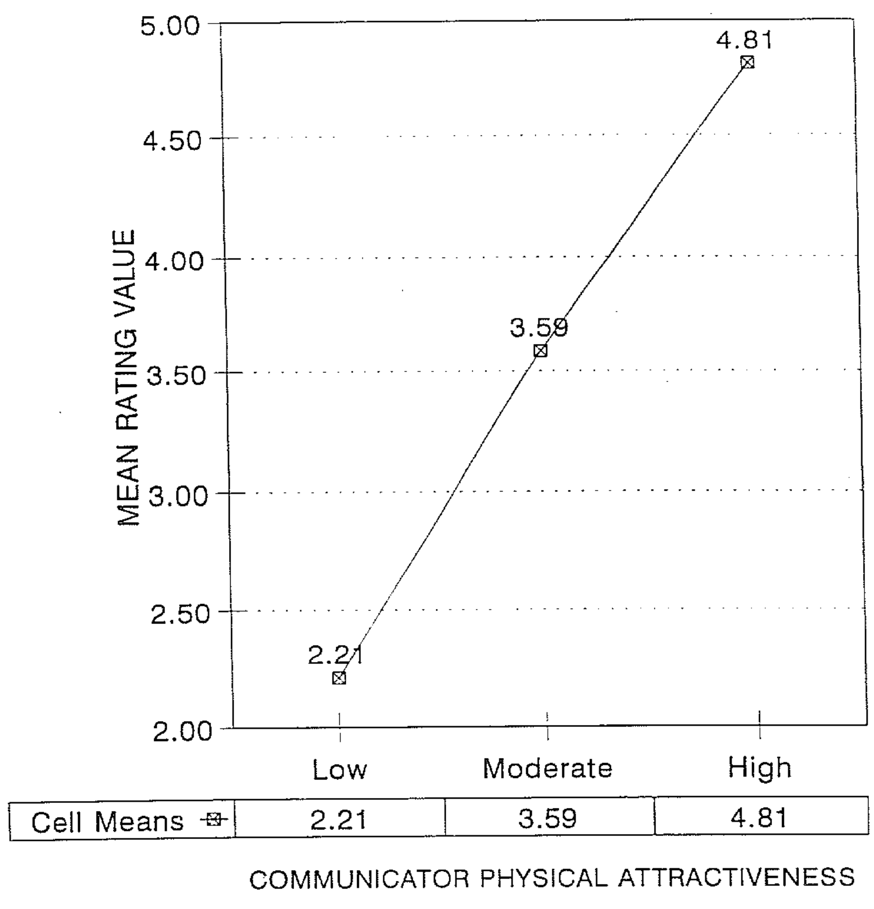
SOURCE OF VARIATION	SS	DF	MS	F	Pr>F
Main Effects	272.91	4	68.23	40.73	0.000
Receiver Sex (A)	1.50	1	1.50	0.90	0.344
Communicator Sex (B)	0.70	1	0.70	0.42	0.517
Communicator Physical Attractiveness (C)	270.70	2	135.35	80.80	0.000
Two-Way Interaction	4.70	5	0.94	0.56	0.729
A x B	1.84	1	1.84	1.10	0.296
A x C	0.23	2	0.11	0.07	0.933
B x C	2.63	2	1.32	0.79	0.457
Three-Way Interaction					
A x B x C	0.10	2	0.05	0.03	0.971
Explained	277.72	11	25.25	15.07	0.000
Residual	381.93	228	1.68		
Total	659.64	239	2.76		

of trustworthiness and expertise scores. Figures 12.2 and 12.3 illustrate these effects of (manipulated) communicator physical attractiveness on perceived communicator trustworthiness and expertise. The Duncan multiple comparison test indicated that all groups in Figure 12.2 (trustworthiness) were different at the .05 level except for the no-photo group and the moderate group. For Figure 12.3 (expertise) both Duncan's and Tukey's tests indicated only the low and high groups were significantly different at the .05 level.

Hypothesis 2: The greater the communicator's (manipulated) physical attractiveness, the greater liking will be expressed for the communicator.

Item number 8 on the third page of the questionnaire booklet asked the subjects what their feelings would be toward the spokesperson for the product's advertising if they had the opportunity to meet the person. Based on the subjects' responses to this question, the second hypothesis was strongly supported: Communicator physical attractiveness had a significant effect on liking for the communicator, while no other main effects or interactions

Figure 12.1
Cell Means of Analysis of Variance Main Effects for the Communicator Physical Attractiveness Manipulation Check



approached significance. Exhibit 12.8 presents the results of the analysis of variance, and Figure 12.4 presents the respective cell mean scores. Multiple comparison tests of the mean scores showed that all groups were different from each other except for the no-photo and moderate groups at the .05 level for both the Duncan and Tukey multiple comparison tests.

> Hypothesis 3: The experimental conditions with communicators of high physical attractiveness will be most effective, as represented by the three attitude components (i.e., affective, cognitive, and conative measures).

Introduction. Items 1 through 9 on page one and items 22 through 24 on page three of the subjects' response booklet were designed to assess the

Exhibit 12.6
Analysis of Variance Using Trustworthiness Ratings as Criterion

SOURCE OF VARIATION	SS	DF	MS	F	Pr>F
Main effects	45.41	5	9.08	6.03	0.000
Receiver Sex (A)	0.11	1	0.11	0.08	0.785
Communicator Sex (B)	2.45	1	2.45	1.63	0.203
Communicator Physical Attractiveness (C)	42.85	3	14.28	9.49	0.000
Two-way interactions	14.95	7	2.14	1.42	0.197
A x B	0.61	1	0.61	0.41	0.524
A x C	7.64	3	2.55	1.70	0.169
B x C	6.70	3	2.33	1.48	0.219
Three-way interaction					
A x B x C	1.14	3	0.40	0. 25	0.860
Explained	61.50	15	4.10	2.72	0.001
Residual	457.68	304	1.51		
Total	519.18	319	1.63		

affective, cognitive, and conative attitude components. While these 12 items were the same as used by an earlier study,[9] the data were subjected to several analyses rather than to blindly use the same component makeup as those authors reported. These analyses were performed in an attempt to ensure that each attitude component was properly defined by the questionnaire items before executing the respective analysis of variances. These preliminary analyses included: (1) Item-to-Total Correlations, (2) Reliability Coefficient Alphas, (3) Factor Analysis, (4) Inter-Item Correlation Coefficients, and (5) Multivariate Analysis of Variance. The three attitude components were believed, a priori, to be measured as follows (note: the corresponding questionnaire item in the subjects' response booklet for each measure is identified in the parenthesis immediately following the measure):

1. Affective Attitude Component—interesting (item 1, p. 1), appealing (item 2, p. 1), impressive (item 4, p. 1), attractive (item 6, p. 1), and eye-catching (item 8, p. 1).
2. Cognitive Attitude Component—believable (item 3, p. 1), informative (item 5, p. 1), and clear (item 7, p. 1).

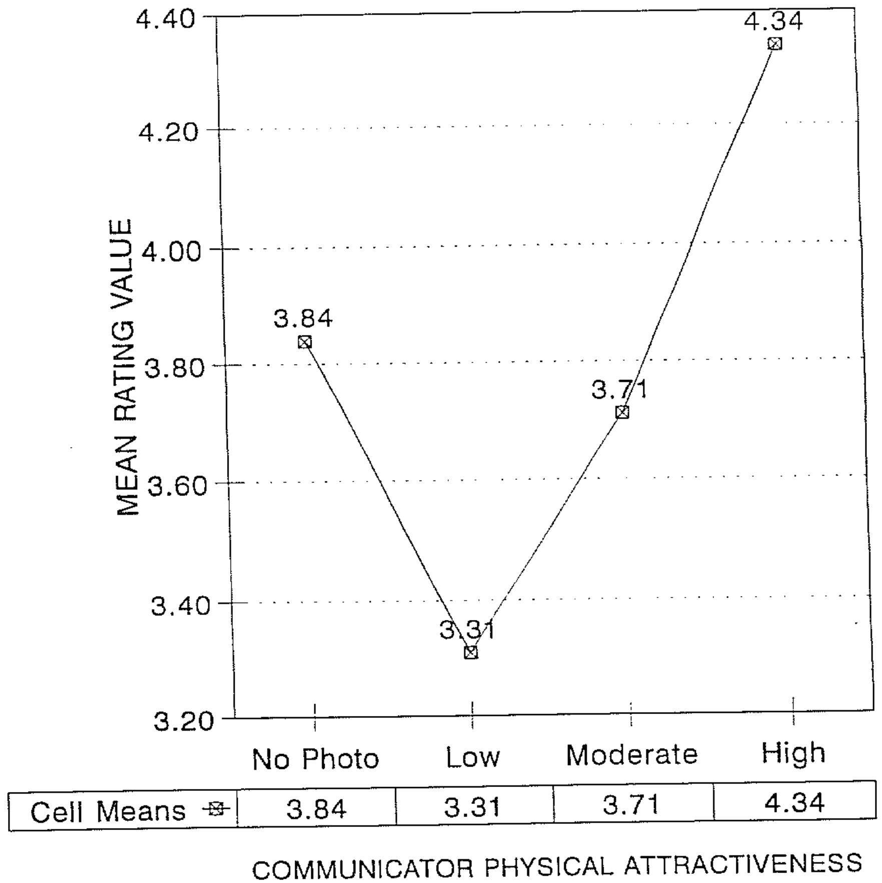
Exhibit 12.7
Analysis of Variance Using Expertise Ratings as Criterion

SOURCE OF VARIATION	SS	DF	MS	F	Pr>F
Main effects	21.67	5	4.33	1.98	0.081
Receiver Sex (A)	0.90	1	0.90	0.41	0.521
Communicator Sex (B)	0.70	1	0.70	0.32	0.571
Communicator Physical Attractiveness (C)	20.06	3	6.69	3.06	0.028
Two-way interactions	16.97	7	2.43	1.11	0.357
A x B	3.40	1	3.40	1.56	0.213
A x C	11.66	3	3.89	1.78	0.151
B x C	1.91	3	0.64	0.29	0.832
Three-way interaction					
A x B x C	10.61	3	3.54	1.62	0.185
Explained	49.25	15	3.28	1.50	0.102
Residual	664.02	304	2.18		
Total	713.27	319	2.24		

3. Conative Attitude Component—seek (item 4, p. 3), try (item 2, p. 3), and buy (item 3, p. 3).

To test these a priori components, all the questionnaire items listed above, as well as the measure of the subjects' overall reaction to the advertising (item 9, p. 1) were subjected to the preliminary analyses.
Item-to-Total Correlations. After the data were collected, the items that a priori defined each of the three components were summed. Based on these sums the item-to-total correlations were calculated for each item for each of the three components. These item-to-total correlations suggested a possible different definition of items for the affective and cognitive components that were thought a priori. Therefore, item-to-total correlations were run again, but with the components changed to represent the suggested definitions. The item-to-total correlations for the a priori definitions and the preliminary data analysis definitions are presented in Exhibit 12.9.
The criteria used to make these judgments were based on the item-to-total correlations of questionnaire items within each of the components. The items then selected, based on the item-to-total correlations, were those

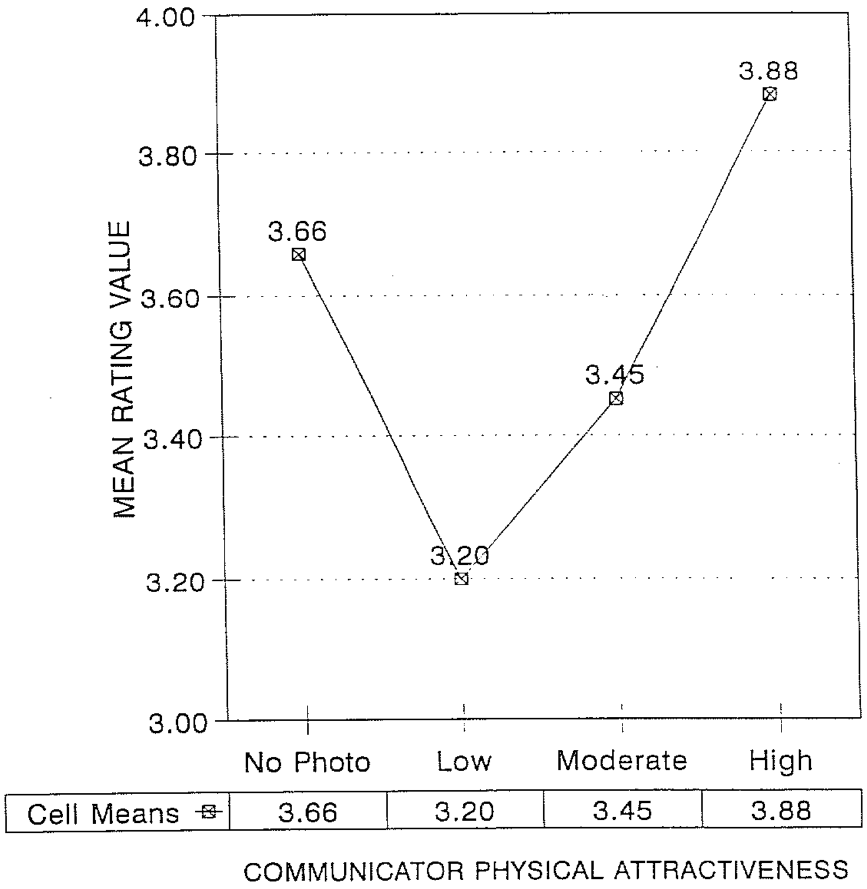
Figure 12.2
Cell Means of Analysis of Variance Main Effects for Physical Attractiveness, Trustworthiness Ratings as Criterion



with the largest values and with a minimum item-to-total correlation value that was not below .50. As illustrated in Exhibit 12.9, these considerations were adhered to in selection of the questionnaire items for the affective, cognitive, and conative attitude components. In addition, "natural groupings" and substantial breaks were observed between the value of items selected and the value of items not selected to represent a specific component.

Reliability Coefficient: Alpha. The alpha reliability coefficient values were in agreement with the other preliminary analyses of the item definitions of each component. The SPSS computer program was used, and two values for alpha were reported for each component. This analysis was performed

Figure 12.3
Cell Means of Analysis of Variance Main Effects for Physical Attractiveness, Expertise Ratings as Criterion



on: (1) the affective component represented by the interesting, appealing, attractive, eye-catching, and overall reaction questionnaire items, (2) the cognitive component represented by the believable, impressive, and informative questionnaire items, and (3) the conative component represented by the try, buy, and seek questionnaire items. For these questionnaire items, the alpha reliability coefficient was .81383 for the affective component, .82782 for the cognitive component, and .86248 for the conative component; the standardized item alpha reliability coefficient values were .93358, .87479, and .94316, respectively. Although no standard of acceptable values appears to exist, the .80+ coefficient values obtained here seem to be exceedingly high in light of the few items making up each component.

Exhibit 12.8
Analysis of Variance Using Liking Ratings as Criterion

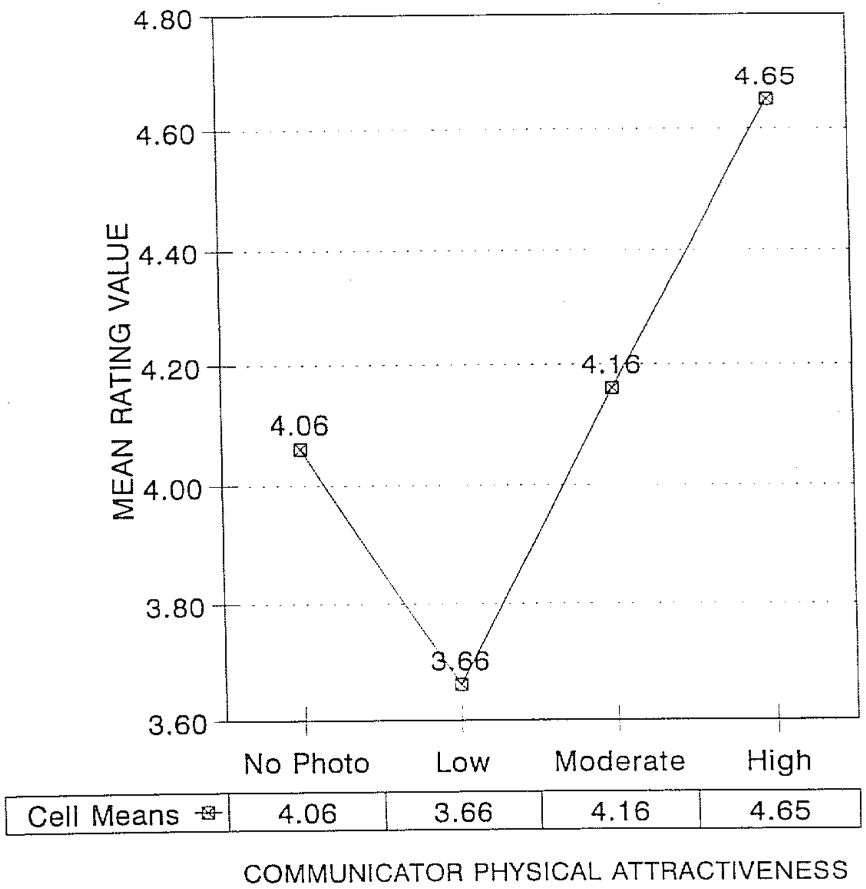
SOURCE OF VARIATION	SS	DF	MS	F	Pr>F
Main effects	43.07	5	8.61	13.28	0.000
Receiver Sex (A)	0.70	1	0.70	1.08	0.299
Communicator Sex (B)	0.70	1	0.70	1.08	0.299
Communicator Physical Attractiveness (C)	41.66	3	13.89	21.41	0.000
Two-way interactions	0.95	7	0.14	0.21	0.983
A x B	0.08	1	0.08	0.12	0.729
A x C	0.68	3	0.23	0.35	0.788
B x C	0.18	3	0.06	0.10	0.963
Three-way interaction					
A x B x C	3.51	3	1.17	1.80	0.146
Explained	47.52	15	3.17	4.89	0.000
Residual	197.15	304	0.65		
Total	244.67	319	0.77		

Factor Analysis. The items used to define each of the three attitude components corresponded with the results from running a factor analysis on the data. Exhibit 12.10 presents the loading on the three factors as obtained from the varimax rotated factor matrix. The eigenvalue of each of these three factors exceeded the minimum eigenvalue of 1.00 which was used throughout this experiment. Specifically, the eigenvalues were 6.01, 1.37, and 1.06 for factors 1, 2, and 3, respectively; while the eigenvalue for a fourth factor was 0.74. The cumulative percentage of variance explained by these three factors was 70.3, represented by 50.1, 11.4, and 8.8 for factors 1, 2, and 3, respectively.

Inter-Item Correlation Coefficients. To further check the validity of each of the three attitude components, correlations were calculated for the items defining each respective component. The resulting correlations were statistically significant for all intracomponent items to at least the .0001 level of probability. These values are presented for the affective component in Exhibit 12.11, the cognitive component in Exhibit 12.12, and the conative component in Exhibit 12.13.

Multivariate Analysis of Variance. Multivariate analysis of variance was

Figure 12.4
Cell Means of Analysis of Variance Main Effects for Physical Attractiveness, Liking Ratings as Criterion



performed on each attitude component. The multivariate analysis indicated statistical significance for each of the components, as well as statistical significance for the univariate analyses for all the items of each component. Although there were interaction effects, the overall main effect of the communicator physical attractiveness was statistically significant for the affective, cognitive, and conative component. These univariate and multivariate statistics are summarized in Exhibit 12.14 for the affective component, Exhibit 12.15 for the cognitive component, and Exhibit 12.16 for the conative component.

Results of Preliminary Analyses. Although generally supportive of the a priori definitions, these preliminary analyses suggested slightly different definitions for the affective and cognitive attitude components. Identification of the questionnaire items used to define each attitude component was based on the unanimous results of the preliminary analyses. The results of these analyses were that the attitude components were best represented by the following questionnaire items:

Exhibit 12.9
Item-to-Total Correlations for Affective, Cognitive, and Conative Attitude Components

QUESTIONNAIRE ITEMS	AFFECTIVE COMPONENT	COGNITIVE COMPONENT	CONATIVE COMPONENT
INTERESTING	(.82)* .82**	(.44) .54	.53
APPEALING	(.88)* .87**	(.44) .54	.51
BELIEVABLE	(.40) .38	(.80)* .82**	.44
IMPRESSIVE	(.76)* .64	(.57) .81**	.50
INFORMATIVE	(.45) .43	(.80)* .78**	.42
ATTRACTIVE	(.85)* .85**	(.44) .54	.47
CLEAR	(.34) .34	(.66)* .38	.21
EYE-CATCHING	(.72)* .76**	(.26) .29	.28
OVERALL REACTION	(.85) .89**	(.60) .67	.58
TRY	(.56) .54	(.49) .57	.89* **
BUY	(.46) .26	(.41) .50	.94* **
SEEK	(.49) .50	(.41) .47	.87* **

(*) Denotes questionnaire item used in a priori definition of the respective attitude component.

(**) Denotes questionnaire item used for definition of respective attitude component after preliminary data analysis.

NOTE:

(1) Values in parentheses are item-to-total correlations based on a priori defined components.

(2) Only one value is listed in the conative component column because the component was not changed, and when the second item-to-total correlations were performed the values remained the same.

1. Affective Attitude Component—interest, appeal, attractive, eye-catching, and overall reaction questionnaire items regarding the advertisement.

Exhibit 12.10
Factor Loadings from the Varimax Rotated Factor Matrix

QUESTIONNAIRE ITEMS	FACTOR 1	FACTOR 2	FACTOR 3
INTERESTING	.63601*	.32165	.28975
APPEALING	.75893*	.29075	.25967
BELIEVABLE	.10798	.27165	.66761*
IMPRESSIVE	.46941	.27080	.54098*
INFORMATIVE	.21638	.25450	.55181*
ATTRACTIVE	.73714*	.23241	.27196
CLEAR	.24492	.03363	.39794
EYE-CATCHING	.67268*	.07395	.09782
OVERALL REACTION	.74084*	.29155	.46009
TRY	.28049	.71220*	.34319
BUY	.15926	.94127*	.17536
SEEK	.28050	.67915*	.21376

(*) Denotes questionnaire items used in definition of the respective attitude component.

NOTE:

Factors 1, 2, and 3 appeared to represent the affective, cognitive, and conative attitudes components, respectively.

2. Cognitive Attitude Component—believable, impressive, and informative questionnaire items regarding the advertisement.
3. Conative Attitude Component—try, buy, and seek questionnaire items regarding the advertised product brand.

The questionnaire item "clear" was not used in the definition of any of the three attitude components.

Affective Component. The affective-component aspect of hypothesis 3 was tested through a procedure that initially averaged the appropriate questionnaire items suggested by the results of the preliminary analyses, that is, interesting, appealing, attractive, eye-catching, and overall reaction. This averaging produced one score for each subject which was then used in the analysis of variance for the affective component. The hypothesis that com-

Exhibit 12.11
Correlation Coefficients Matrix for the Affective Attitude Component Items

	APPEALING	ATTRACTIVE	EYE-CATCHING	OVERALL REACTION
INTERESTING	.77157 .0001	.59393 .0001	.45548 .0001	.72230 .0001
APPEALING		.69983 .0001	.52441 .0001	.72848 .0001
ATTRACTIVE			.58862 .0001	.74664 .0001
EYE-CATCHING				.60882 .0001

NOTE: Correlation Coefficients / Probability under Ho=0

municator physical attractiveness would have a positive influence upon the affective component was confirmed at the .000 level of probability (Exhibit 12.17), and is illustrated in Figure 12.5. Duncan's multiple comparison test indicated that all four groups were significantly different from each other at the .05 level for the main effect of communicator physical attractiveness.

In summary, hypothesis 3 was accepted. This acceptance was based on the data analyses that showed greater communicator physical attractiveness resulted in greater values for the affective, cognitive, and conative attitude components.

Cognitive Component. The cognitive-component aspect of hypothesis 3 was tested through a procedure that initially averaged the appropriate questionnaire items suggested by the results of the preliminary analyses, that is, believable, impressive, and informative. This averaging produced one score for each subject which was then used in the analysis of variance for the cognitive component. The hypothesis that the communicator physical attractiveness would have a positive influence upon the cognitive component was confirmed at the .000 level of significance (Exhibit 12.18), and is illustrated in Figure 12.6. Duncan's multiple comparison test indicated that the only significant difference at the .05 level for the main effect of communicator physical attractiveness was between the high physical attractiveness group and each of the other three groups.

Conative Component. The final use of these preliminary analyses for hypothesis 3 was for the conative attitude component. The initial step was to average the appropriate questionnaire items suggested by the results of the preliminary analyses, that is, try, buy, and seek. This averaging produced one score for each subject which was then used in the analysis of

Exhibit 12.12
Correlation Coefficients Matrix for the Cognitive Attitude Component Items

	IMPRESSIVE	INFORMATIVE
BELIEVABLE	.50369 .0001	.49172 .0001
IMPRESSIVE		.45036 .0001

NOTE: Correlation Coefficients / Probability under Ho=0

variance for the conative component. The hypothesis that the communicator physical attractiveness would have a positive influence upon the conative component was supported at the .000 level of significance (Exhibit 12.19), and is illustrated in Figure 12.7. Duncan's multiple comparison test indicated that for the main effect of communicator physical attractiveness, the no-photo group and moderate group were not significantly different at the .05 level, but that each of these groups was significantly different from both the low and high groups. Similarly, the low and high groups were significantly different from each other for the main effect of communicator physical attractiveness.

Hypothesis 4: The receivers exposed to the persuasive communication with communicators of higher physical attractiveness will respond with more positive and stronger beliefs than receivers exposed to communicators of lower physical attractiveness.

Preliminary Analyses. Items 1 to 7 and item 11, on the second page of the subjects' response booklet were designed to assess the influence of the experimental treatments on the subjects' beliefs. These items dealt with subjects' beliefs about attributes the advertising claimed the product possessed. Before performing the appropriate analysis of variance on the subjects' belief responses, a belief component was identified through several preliminary analyses of the data.

The data subjected to these preliminary analyses involved six belief items that were specifically addressed in the advertisement mock-up, and five belief items that were not addressed in the advertisement mock-up. The addressed product: (1) provided relief from headaches, (2) provided relief from minor pains, (3) was effective for both males and females, (4) was safe, (5) was strong, and (6) worked fast. The questionnaire items not addressed dealt with beliefs about the advertised product's (1) price, (2) quality, (3) uniqueness, (4) taste, and (5) the American Medical Association's seal of approval.

Exhibit 12.13
Correlation Coefficients Matrix for the Conative Attitude Component Items

	BUY	SEEK
TRY	.77632 .0001	.62842 .0001
BUY		.71202 .0001

NOTE: Correlation Coefficients / Probability under Ho=0

Because there was no a priori definition of the belief component to be used in the analysis of variance, factor analysis was first performed on all 11 belief questionnaire items. The varimax rotated matrix (Exhibit 12.20) revealed the factor loading appropriate for the belief components were the questionnaire items regarding: (1) provides headache relief, (2) provides relief from minor pain, (3) effectiveness for both sexes, (4) fast, (5) strong, and (6) safe. The results of this initial factor analysis indicated those questionnaire belief items that were not addressed in the advertisement mock-up should be excluded from the belief component (Exhibit 12.20). The criterion used for this factor analysis was a minimum eigenvalue of 1.00 for each factor. Specifically, the eigenvalues were 4.32 and 1.49 for factors 1 and 2, respectively. The cumulative percentage of variance explained was 52.9, represented by 39.3 for factor 1 and 13.6 for factor 2.

Furthermore, this initial factor analysis suggested that those questionnaire items not addressed in the advertisement mock-up and which were used for hypothesis 5 (product price, quality, and uniqueness), should also be excluded from the belief component (Exhibit 12.20). These items, except for the AMA seal of approval item, were excluded from a second factor analysis that resulted in one factor (Exhibit 12.21). As shown in Exhibit 12.21, this second factor analysis clarified the appropriate questionnaire items for the belief component to be: (1) provides headache relief, (2) provides relief from minor pain, (3) effectiveness for both sexes, (4) fast, and (5) strong. Again, the criterion of a minimum eigenvalue of 1.00 was followed. Specifically, the eigenvalue for this factor was 3.73 while the percentage of variance explained was 53.3. The eigenvalue and percentage of variance explained by a second factor was 0.88 and 12.5, respectively.

The definition of items for the belief component suggested by factor analysis was supported by four additional analyses. First, the item-to-total correlations (Exhibit 12.22) revealed high correlation values for all items

Exhibit 12.14
Univariate and Multivariate Analysis of Variance of the Affective Attitude Component

	DEPENDENT VARIABLES					
	INT.	APP.	ATT.	EYE.	OVER.	MULTI-VARIATE
A	0.34 .5586	1.50 .2211	2.54 .1123	8.17 .0046	6.81 .0095	2.36 .0402
B	2.80 .0953	1.08 .3003	2.83 .0933	0.34 .5581	1.35 .2469	0.88 .4984
C	25.79 .0001	37.67 .0001	42.85 .0001	21.25 .0001	34.52 .0001	10.09 .0001
A x B	0.17 .6760	1.28 .2586	0.17 .6825	0.34 .5581	1.58 .2098	1.67 .1408
A x C	10.87 .0001	10.02 .0001	9.75 .0001	5.06 .0021	12.12 .0001	3.49 .0001
B x C	1.53 .2063	3.70 .0122	4.18 .0065	1.56 .1982	3.41 .0177	2.09 .0087
AxBxC	6.99 .0002	9.40 .0001	4.16 .0067	1.77 .1507	7.41 .0001	2.67 .0006

NOTE: (1) F Value / Probability > F

(2) Dependent Variables (df=319):
"INT." represents interesting (MS=16.53)
"APP." represents appealing (MS=17.44)
"ATT." represents attractive (MS=18.95)
"EYE." represents eye-catching (MS=15.17)
"OVER." represents overall reaction (MS=16.24)

(3) Source of Variation:
"A" is receiver sex
"B" is communicator sex
"C" is communicator physical attractiveness

(4) "MULTIVARIATE" represents multivariate analysis of variance with (Wilk's Criterion)

making up the belief component. Second, the reliability coefficients alpha and standardized item alpha were .80555 and .91136, respectively, for the belief component. Third, items within the belief component were all highly correlated to a statistical significance level of at least .0001 (Exhibit 12.23). Finally, the multivariate analysis of variance revealed statistical significance for an overall communicator's physical attractiveness effect on the belief component at the .0084 level (Exhibit 12.24).

The criterion used in regard to the item-to-total correlations was to select

Exhibit 12.15
Univariate and Multivariate Analysis of Variance of the Cognitive Attitude Component

	DEPENDENT VARIABLES			
	BELIEVE	IMPRESS	INFORM	MULTI-VARIATE
A	0.10 .7574	3.52 .0615	0.18 .6728	1.27 .2852
B	0.38 .5369	3.52 .0615	0.18 .6728	2.13 .0947
C	14.40 .0001	20.96 .0001	9.65 .0001	8.73 .0001
A x B	0.86 .3545	0.67 .4151	1.21 .2726	0.52 .6744
A x C	0.59 .6266	9.01 .0001	2.81 .0389	3.77 .0001
B x C	1.78 .1485	3.75 .0114	1.02 .3841	1.43 .1712
AxBxC	1.45 .2272	6.57 .0003	2.42 .0648	2.68 .0046

NOTE: (1) F Value / Probability > F

(2) Dependent Variables (df=319)
"BELIEVE" represents believable (MS=7.82)
"IMPRESS" represents impressive (MS-16.09)
"INFORM" represents informative (MS=5.75)

(3) Source of Variation:
"A" is receiver sex
"B" is communicator sex
"C" is communicator physical attractiveness

(4) "MULTIVARIATE represents multivariate analysis of variance (Wilk's Criterion)

the highest values once other minimum requirements were exceeded. The criterion used in regard to determining the sufficient coefficient alpha values was based on the notion that a value of .80+ is exceedingly high for a component with only a few questionnaire items.

Analysis of Variance. The belief component was tested through a procedure that initially averaged the appropriate questionnaire items suggested in the results of the preceding "Preliminary Analyses" section. These questionnaire items were: (1) product provides relief from headache, (2) product provides relief from minor pain, (3) product is effective for both sexes, (4) product is strong, and (5) product is fast.

Exhibit 12.16
Univariate and Multivariate Analysis of Variance of the Conative Attitude Component

	DEPENDENT VARIABLES			
	TRY	BUY	SEEK	MULTI-VARIATE
A	3.03 .0829	0.79 .3747	0.08 .7827	1.23 .2980
B	2.48 .1165	0.00 .9455	0.00 .9686	1.70 .1642
C	15.59 .0001	13.16 .0001	23.07 .0001	8.59 .0001
A x B	4.29 .0392	2.06 .1520	0.35 .5545	1.57 .1956
A x C	9.15 .0001	4.07 .0076	3.42 .0177	3.20 .0009
B x C	3.18 .0241	0.86 .4668	2.88 .0354	1.87 .0532
AxBxC	5.19 .0018	2.90 .0347	3.57 .0146	2.19 .0208

NOTE: (1) F Value / Probability > F

(2) Dependent Variables (df=319)
TRY (MS=13.29)
BUY (MS=11.73)
SEEK (MS=13.29)

(3) Source of Variation:
"A" is receiver sex
"B" is communicator sex
"C" is communicator physical attractiveness

(4) "MULTIVARIATE" represents multivariate analysis of variance (Wilk's Criterion)

The results of the data analysis supported the fourth hypothesis: Communicator physical attractiveness had a significant positive effect on the receivers' beliefs. Exhibit 12.25 presents the summary of the analysis of variance, and Figure 12.8 illustrates the respective mean scores for the main effects of communicator physical attractiveness which were significant at the .003 level. For the significant main effect of communicator physical attractiveness, Duncan's multiple comparison test indicated that the low group was different than both the moderate and high groups, and that the no-photo group was different than the high group. These differences were significant at the .05 level of probability.

Exhibit 12.17
Analysis of Variance Using the Affective Attitude Component Ratings as Criterion

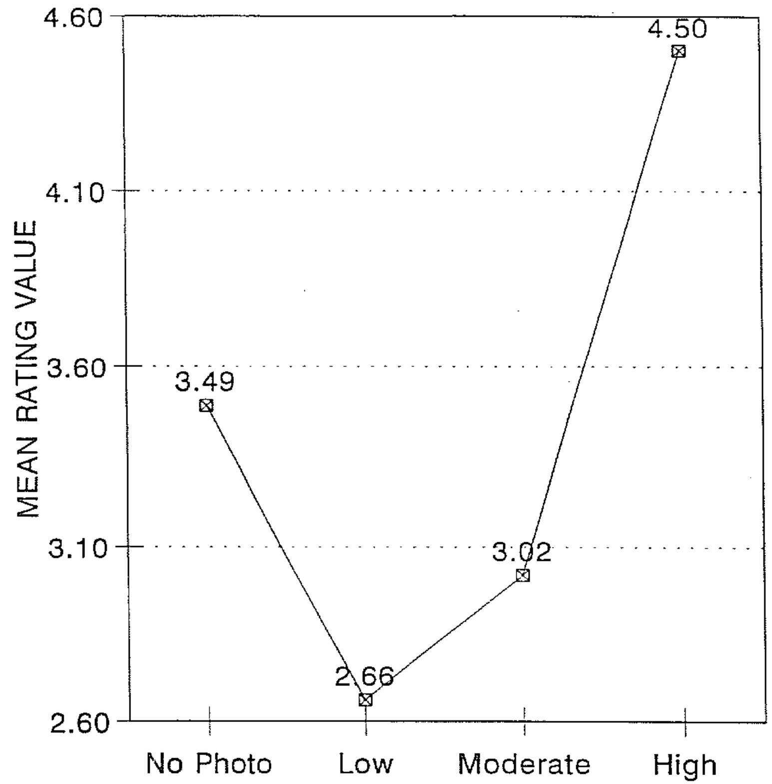
SOURCE OF VARIATION	SS	DF	MS	F	Pr>F
Main effects	160.77	5	32.16	32.73	0.000
Receiver Sex (A)	5.36	1	5.36	5.46	0.020
Communicator Sex (B)	2.42	1	2.42	2.46	0.118
Communicator Physical Attractiveness (C)	153.00	3	51.00	51.91	0.000
Two-way interactions	53.57	7	7.65	7.79	0.000
A x B	0.03	1	0.03	0.03	0.867
A x C	44.60	3	14.87	15.13	0.000
B x C	8.94	3	2.98	3.03	0.030
Three-way interaction					
A x B x C	25.53	3	8.51	8.66	0.000
Explained	239.88	15	15.99	16.28	0.000
Residual	298.68	304	0.98		
Total	538.55	319	1.69		

Hypothesis 5: Recall of advertisement details or facts will be less with communicators of higher physical attractiveness than with communicators of lower physical attractiveness.

This fifth hypothesis was not supported by the results: Recall of advertisement details or facts, as manipulated in this experiment, was not significantly affected by communicator physical attractiveness. Questionnaire items 1 through 7 on the fourth page of the subjects' response booklet were multiple choice questions about specific details presented in the message of the advertisement mock-up. These seven questions were summed to produce a single score for each subject. This single score was then used in the analysis of variance to test the differences of recall due to the experimental treatments.

The analysis of variance revealed no main effects or three-way interaction effects, but did indicate one two-way interaction between receiver sex and communicator sex. Exhibit 12.26 presents the results of this analysis of variance, and Figure 12.9 illustrates the two-way interaction by presenting the respective mean scores for recall. The interaction was due to higher

Figure 12.5
Cell Means of Analysis of Variance Main Effects for Physical Attractiveness, Affective Component Ratings as Criterion



	No Photo	Low	Moderate	High
Cell Means	3.49	2.66	3.02	4.50

COMMUNICATOR PHYSICAL ATTRACTIVENESS

recall when male and female receivers were confronted with a communicator of the opposite sex, rather than when confronted by a same-sex communicator.

> Hypothesis 6: Perceptions of product quality, product price, and product uniqueness will be positively correlated with perceived communicator physical attractiveness.

Items 7, 8, and 9 on the second page of the subjects' response booklet asked the subjects to indicate their beliefs about the price, quality, and uniqueness of the advertised brand in relation to other major brands. Item 7 on the third page of the subjects' response booklet asked them to indicate

Exhibit 12.18
Analysis of Variance Using the Cognitive Attitude Component Ratings as Criterion

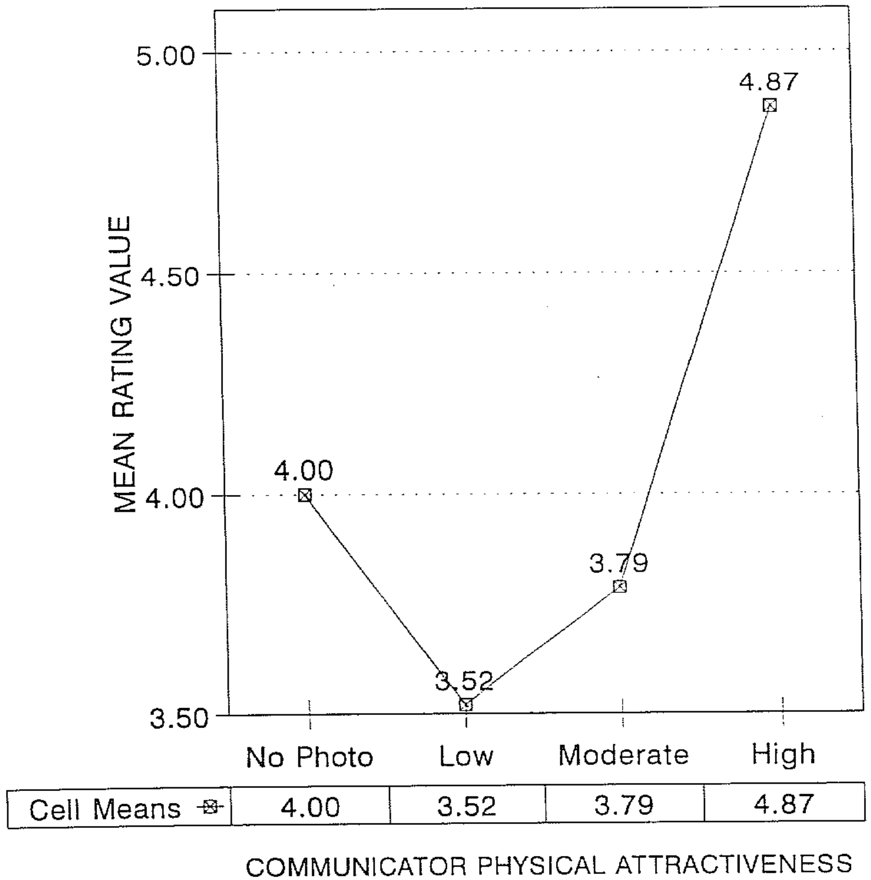
SOURCE OF VARIATION	SS	DF	MS	F	Pr>F
Main effects	82.88	5	16.58	14.48	0.000
Receiver Sex (A)	1.42	1	1.42	1.24	0.267
Communicator Sex (B)	0.14	1	0.14	0.12	0.727
Communicator Physical Attractiveness (C)	81.32	3	27.11	23.68	0.000
Two-way interactions	28.97	7	4.14	3.62	0.001
A x B	1.71	1	1.71	1.49	0.223
A x C	15.97	3	5.32	4.65	0.003
B x C	11.29	3	3.77	3.29	0.021
Three-way interaction					
A x B x C	15.81	3	5.27	4.61	0.004
Explained	127.66	15	8.51	7.44	0.000
Residual	347.93	304	1.15		
Total	475.59	319	1.49		

their evaluation of the physical attractiveness of the spokesperson for the brand's advertising. The individual scores for each of these four questionnaire items were used to calculate correlation values for perceptions of communicator physical attractiveness with perceptions of each of these three product characteristics.

This sixth hypothesis was not supported: Perceived product price (Pearson r = .0088; significance = .434) and perceived product uniqueness (Pearson r = .0350; significance = .282) were not positively correlated with perceived communicator physical attractiveness. Although perceived product quality was significantly correlated with perceived communicator physical attractiveness, the correlation was low. The weak, but statistically significant, correlation that did emerge was a correlation of .1084 (Pearson r) between perceived physical attractiveness and perceived product quality, which was significant at the .020 level.

> Hypothesis 7: The effects predicted in hypotheses 1 through 5 will be monotonic; that is, persuasive communication using communicators of high physical attractiveness will be most effective followed by communicators of

Figure 12.6
Cell Means of Analysis of Variance Main Effects for Physical Attractiveness, Cognitive Component Ratings as Criterion



moderate physical attractiveness, and least effective will be communicators of low physical attractiveness.

The seventh hypothesis was confirmed: The relationship between communicator physical attractiveness and its effects was monotonic in those hypotheses which exhibited significant effects. These communicator physical attractiveness effects for hypothesis 1 are illustrated for perceived trustworthiness in Figure 12.2 and for perceived expertise in Figure 12.3. The monotonic liking effects due to communicator physical attractiveness are displayed in Figure 12.4. The hypothesized monotonic effects for hypothesis 3 are presented for the affective attitude component in Figure 12.5, for

Exhibit 12.19
Analysis of Variance Using the Conative Attitude Component Ratings as Criterion

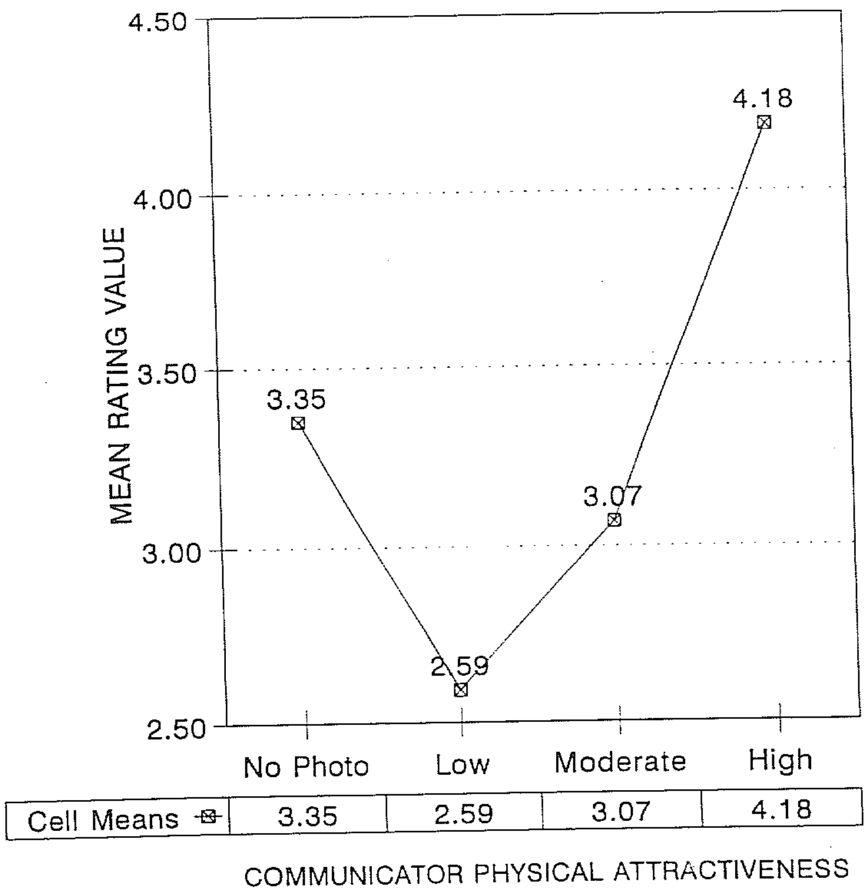
SOURCE OF VARIATION	SS	DF	MS	F	Pr>F
Main effects	109.44	5	21.89	13.28	0.000
Receiver Sex (A)	1.96	1	1.96	1.19	0.277
Communicator Sex (B)	0.53	1	0.53	0.32	0.571
Communicator Physical Attractiveness (C)	106.96	3	35.65	21.64	0.000
Two-way interactions	49.24	7	7.03	4.27	0.000
A X B	3.98	1	3.98	2.42	0.121
A X C	32.82	3	10.94	6.64	0.000
B X C	12.44	3	4.15	2.52	0.058
Three-way interaction					
A X C	23.34	3	7.78	4.72	0.003
Explained	182.02	15	12.14	7.36	0.000
Residual	500.93	304	1.65		
Total	682.95	319	2.14		

the cognitive attitude component in Figure 12.6, and for the conative attitude component in Figure 12.7. This relationship was evidenced again, in Figure 12.8, for the fourth hypothesis which dealt with beliefs. Finally, the fifth hypothesis cannot be cited in either support or nonsupport of a monotonic relationship because, as shown in Exhibit 12.26, there was no significant effect due to communicator physical attractiveness.

SUMMARY

The underlying motivation behind this experiment was to answer theoretically based, empirical questions regarding the effects of communicator physical attractiveness upon persuasive communication effectiveness. The physical attractiveness variable was manipulated through photographs of individuals whose physical attractiveness was determined by a consensus of judges through two different rating procedures at two different times for each procedure. Different subjects were used for each method and no subjects used in these initial ratings were used in later phases of this experiment.

Figure 12.7
Cell Means of Analysis of Variance Main Effects for Physical Attractiveness, Conative Component Ratings as Criterion



The dependent variable, persuasive communication effectiveness, was measured by administering posttreatment questionnaires. These questionnaires assessed beliefs about the product, attitudes toward the product, evaluations of the advertising, recall of details stated in the message, and perceptions of the communicator. The laboratory experimental procedure involved a $2 \times 2 \times 4$ factorial design. The independent variables were the receivers' sex (male and female), the communicators' sex (male and female), and the communicators' physical attractiveness (low, moderate, and high physical attractiveness, as well as a control condition with no information). The dependent variables were subjects' responses to the persuasive com-

Exhibit 12.20
Factor Loadings Obtained from the Varimax Rotated Factor Matrix

	Factor 1	Factor 2
PROVIDES HEADACHE RELIEF	.81302*	.09129
PROVIDES RELIEF FROM MINOR PAIN	.73279*	.10860
EFFECTIVE FOR BOTH SEXES	.72107*	.09803
SAFE	.59415*	.10637
STRONG	.59094*	.41273
FAST	.70261*	.34132
PRICE RELATION	.13733	.54514*
QUALITY RELATION	.30965	.66871*
UNIQUENESS RELATION	.18181	.61319*
TASTE	-.00101	.30511
AMA SEAL OF APPROVAL	.33428	.18250

(*) Denotes relatively high factor loading values on that factor.

(**) Denotes moderately high factor loading values on that factor.

munication measures mentioned above. Persuasive communication effectiveness was defined as the difference in responses given by subjects who received different experimental treatments (i.e., persuasive communication with communicators of low, moderate, and high physical attractiveness and a control group with no physical attractiveness information). These responses were in regard to a variety of hypothesized effects (e.g., perceived trustworthiness, expertise, and liking, attitudinal components, beliefs, and recall of details). Communicator physical attractiveness was operationally defined through use of a truth-of-consensus method performed in phase one.

Procedures and Analyses

The methodology was divided into two phases and involved a total of 602 different subjects. The first phase identified the physical attractiveness

Exhibit 12.21
Factor Loadings Obtained from the Principle Factor with Iterations

	FACTOR 1
PROVIDES HEADACHE RELIEF	.79165*
PROVIDES RELIEF FROM MINOR PAIN	.72775*
EFFECTIVE FOR BOTH SEXES	.71099*
SAFE	.60724
STRONG	.68136*
FAST	.78123*
AMA SEAL OF APPROVAL	.37431

(*) Denotes factor loading values of items selected for the belief component.

of stimulus persons to be used as communicators of persuasive communication in the second phase. The second phase was the 2 × 2 × 4 factorial experiment.

Although analysis of variance was the predominant analytic technique, a number of other procedures were also employed. Analysis of variance was the exclusive test for hypothesis 1 (regarding trustworthiness and expertise), hypothesis 2 (regarding liking), and hypothesis 5 (regarding recall). Hypothesis 3 (regarding the affective, cognitive, and conative attitude components) and hypothesis 4 (regarding the belief component) were, also, ultimately tested through analysis of variance. However, for these two hypotheses, the data were subjected to a number of other statistical techniques in preparation for the analysis of variance tests. These other techniques included: (1) item-to-total correlations, (2) alpha reliability coefficients, (3) factor analysis, (4) intra-component item correlations, and (5) multivariate analysis of variance. For hypothesis 6 (regarding product price, quality, and uniqueness) Pearson Product Moment Correlations were performed. Finally, for hypothesis 7 (regarding the monotonic relationship) observation of the relationships between communicator physical attractiveness and its hypothesized effects, as exhibited in hypotheses 1 through 5, was used. This observation involved the figures illustrating the analysis of variance mean effects due to communicator physical attractiveness for each of the respective hypotheses.

Exhibit 12.22
Item-to-Total Correlations for the Belief Component Items

	BELIEF COMPONENT
BELIEF COMPONENT	1.00000
PROVIDES HEADACHE RELIEF	.83009*
PROVIDES RELIEF FROM MINOR PAIN	.79761*
EFFECTIVE FOR BOTH SEXES	.75437*
SAFE	.54948
STRONG	.78036*
FAST	.83044*
PRICE RELATION	.26955
QUALITY RELATION	.44624
UNIQUENESS RELATION	.34102
TASTE	.07378
AMA SEAL OF APPROVAL	.33206

(*) Denotes questionnaire items used for definition of the belief component after the preliminary data analyses.

Findings

Overall, this experiment suggests that communicator physical attractiveness does have an effect on persuasive communication effectiveness. Furthermore, consistent with theoretical prediction and explanation, this effect seems to be monotonic, that is, as communicator physical attractiveness increases, persuasive communication effectiveness also increases.

The major findings are:

1. A monotonic relationship existed between (manipulated) communicator physical attractiveness and perceptions of communicator trustworthiness, expertise, and attraction (liking), regardless of communicator sex and/or receiver sex.
2. Although some interaction effects occurred, the relationship between the effects of communicator physical attractiveness and the affective, cognitive, and conative attitudinal components was positive.
3. Recall of advertisement detail was not substantially different for different levels of communicator physical attractiveness.

Exhibit 12.23
Correlation Coefficients Matrix for the Belief Component Items

	MINOR PAIN RELIEF	EFFECTIVE FOR BOTH SEXES	STRONG	FAST
PROVIDES HEADACHE RELIEF	.6878 .0001	.5911 .0001	.4967 .0001	.5985 .0001
PROVIDES RELIEF FROM MINOR PAIN		.5530 .0001	.4428 .0001	.5199 .0001
EFFECTIVE FOR BOTH SEXES			.4397 .0001	.5157 .0001
STRONG				.6854 .0001

NOTE: Correlation Coefficients / Probability under Ho=0

4. Perceptions of product price, quality, and uniqueness did not correlate with perceptions of communicator physical attractiveness.
5. The influence of communicator physical attractiveness upon beliefs was monotonic, that is, as the consumer physical attractiveness increased, the receivers' stated beliefs increased in agreement with the persuasive communication.
6. Low, moderate, and high communicator physical attractiveness produced negative, neutral, and positive effects on persuasive communication effectiveness, respectively. This observation is arrived at by first using the data obtained from the experimental control condition, of no physical attractiveness information (i.e., no photograph), as a standard of comparison or a norm. Then, the collected data for each of the experimental conditions of low, moderate, and high physical attractiveness are compared against this norm.

Exhibit 12.24
Univariate and Multivariate Analysis of Variance of the Belief Component

	DEPENDENT VARIABLES					
	HEAD	PAIN	SEX	STRG	FAST	MULTI-VARIATE
A	0.02 .8783	1.53 .2163	1.35 .2462	0.52 .4728	0.02 .8848	1.04 .3966
B	3.57 .0597	1.79 .1819	0.71 .3488	0.03 .8575	1.03 .3110	1.10 .3619
C	2.67 .0467	2.37 .0697	1.76 .1540	5.50 .0012	5.11 .0020	2.10 .0084
A x B	3.97 .0473	4.97 .0265	3.22 .0735	0.98 .3238	0.02 .8848	2.10 .0646
A x C	1.53 .2063	1.58 .1924	0.87 .4571	2.01 .1112	2.79 .0403	1.74 .1267
B x C	1.53 .2062	2.90 .0349	1.46 .2247	2.60 .0514	2.03 .1085	1.43 .1267
AxBxC	3.95 .0088	3.68 .0125	1.94 .1206	1.93 .1230	1.81 .1426	1.52 .0926

NOTE: (1) F Value / Probability > F

(2) Dependent Variables (df=319):
"HEAD"--provides headache relief (MS=2.92)
"PAIN"--provides minor pain relief (MS=3.38)
"SEX"--effective for both sexes (MS=1.75)
"STRG"--strong (MS=3.89)
"FAST"--works fast (MS=3.23)

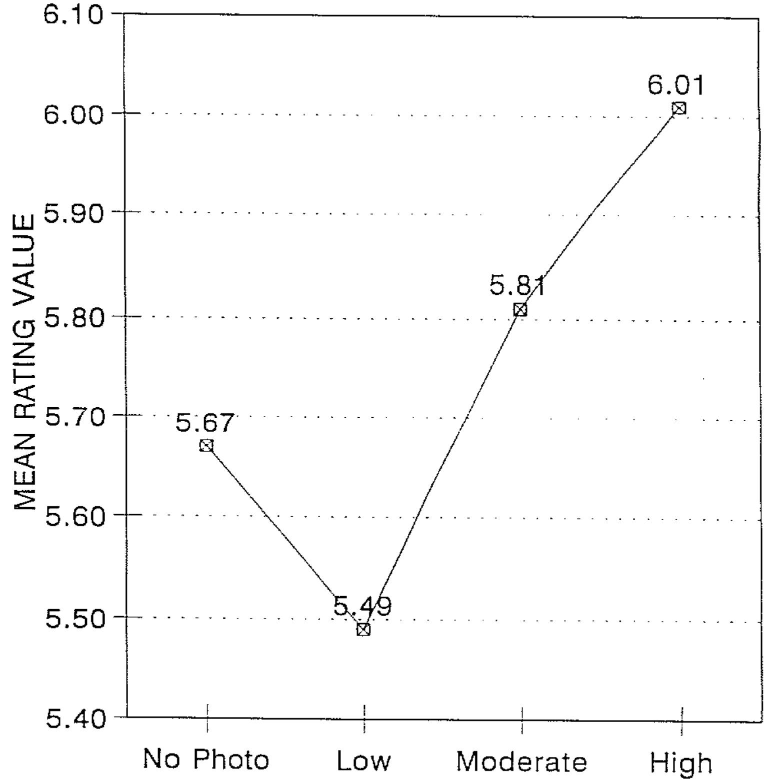
(3) Source of Variation:
"A" is receiver sex
"B" is communicator sex
"C" is communicator physical attractiveness

(4) "MULTIVARIATE" represents multivariate analysis of variance (Wilk's Criterion)

Exhibit 12.25
Analysis of Variance Using Belief Component Ratings as Criterion

SOURCE OF VARIATION	SS	DF	MS	F	Pr>F
Main effects	13.06	5	2.61	3.23	0.007
Receiver Sex (A)	0.17	1	0.17	0.21	0.647
Communicator Sex (B)	1.17	1	1.17	1.45	0.230
Communicator Physical Attractiveness (C)	11.71	3	3.90	4.82	0.003
Two-way interactions	13.22	7	1.89	2.33	0.025
A x B	2.35	1	2.35	2.90	0.089
A x C	4.26	3	1.42	1.75	0.156
B x C	6.61	3	2.20	2.72	0.044
Three-way interaction					
A x B x C	8.73	3	2.91	3.60	0.014
Explained	35.01	15	2.33	2.88	0.000
Residual	246.08	304	0.81		
Total	281.09	319	0.88		

Figure 12.8
Cell Means of Analysis of Variance Main Effects for Physical Attractiveness, Belief Component Ratings as Criterion



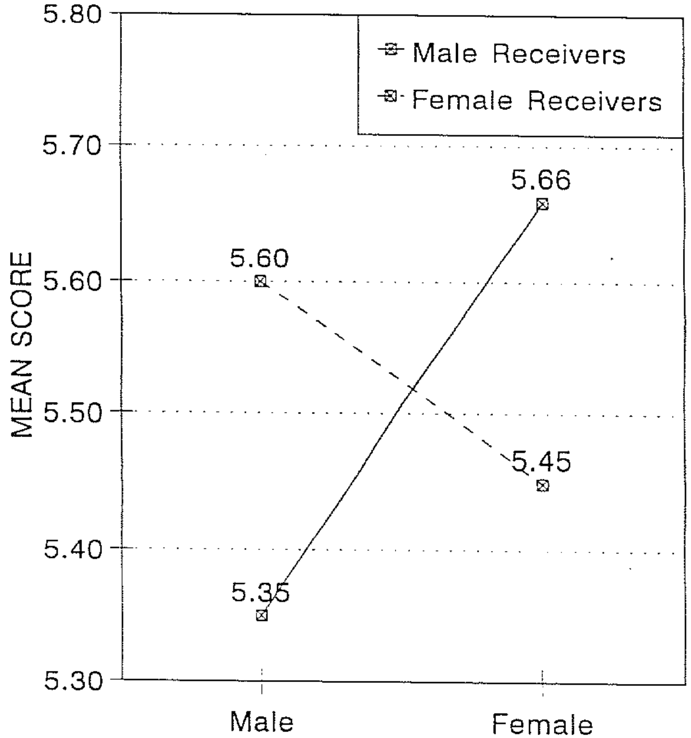
Cell Means	5.67	5.49	5.81	6.01

COMMUNICATOR PHYSICAL ATTRACTIVENESS

Exhibit 12.26
Analysis of Variance Using Recall Scores as Criterion

SOURCE OF VARIATION	SS	DF	MS	F	Pr>F
Main effects	3.04	5	0.61	0.60	0.699
Receiver Sex (A)	0.03	1	0.03	0.03	0.868
Communicator Sex (B)	0.53	1	0.53	0.52	0.471
Communicator Physical Attractiveness (C)	2.48	3	0.83	0.82	0.485
Two-way interactions	8.80	7	1.26	1.24	0.280
A x B	4.28	1	4.28	4.27	0.041
A x C	1.33	3	0.45	0.44	0.725
B x C	3.18	3	1.06	1.05	0.371
Three-way interaction					
A x B x C	2.33	3	0.78	0.77	0.512
Explained	14.17	15	0.95	0.93	0.527
Residual	307.74	304	1.01		
Total	321.91	319	1.01		

Figure 12.9
Cell Means of Analysis of Variance Two-Way Interaction Effects, Recall Scores as Criterion



	Male	Female
Male Receivers	5.35	5.66
Female Receivers	5.60	5.45

COMMUNICATOR GENDER

NOTES

1. Readers interested in this supporting research literature are referred to Gordon L. Patzer, *The Physical Attractiveness Phenomena* (New York: Plenum Publishing Corporation, 1985); and Gordon L. Patzer, *An Experimental Investigation of the Relationship between Communication Facial Physical Attractiveness and Non-Personal Persuasive Communication Effectiveness in Marketing* (Ph.D. diss., Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1980).
2. D. R. Campbell and D. W. Fiske, "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin* 56 (1959), pp. 81–105.
3. For example, see: G. A. Churchill, Jr., "A Paradigm for Developing Better

Measures of Marketing Constructs," *Journal of Marketing Research* 16 (1979), pp. 64–73.

4. S. Freud, in *Freudian Dictionary of Psychoanalysis*, trans. N. Foder and F. Gaynor (Westport, Conn.: Greenwood Press, 1958).

5. Note that Nunnally (p. 226) proposes that for basic research, reliability values of .50 and .60 are sufficient, and that attempts to obtain values above .80 are a waste of effort. J. C. Nunnally, *Psychometric Theory* (New York: McGraw-Hill, 1967).

6. M. Snyder and M. Rothbart, "Communicator Attractiveness and Opinion Change," *Canadian Journal of the Behavioural Sciences* 3 (1971), pp. 377–387.

7. W. Oaks, "External Validity and the Use of Real People as Subjects," *American Psychologist* 27 (1972), pp. 959–962.

8. M. B. Holbrook and J. A. Howard, "Frequently Purchased Nondurable Goods and Services," in *Selected Aspects of Consumer Behavior*, R. Ferber, ed. (Washington, D.C.: National Science Foundation, 1977).

9. M. J. Baker and G. A. Churchill, Jr., "The Impact of Physically Attractive Models on Advertising Evaluations," *Journal of Marketing Research* 14 (1977), pp. 538–555.

Concluding Comment

The author of this book has found experiments to be a robust and productive methodology for marketing research both in the United States and, literally, around the world. He continues to work with experiments in marketing research and welcomes comments and questions about this book and related topics. Please address correspondence to:

Gordon L. Patzer, Ph.D.
Dean and Professor
School of Business Administration
California State University, Stanislaus
801 West Monte Vista Avenue
Turlock, California 95380

Selected Bibliography

Carson, Richard T., Jordan J. Louviere, Don A. Anderson, Phipps Arabie, David S. Bunch, David A. Hensher, Richard M. Johnson, Warren F. Kuhfeld, Dan Steinberg, Joffre Sait, Harry Timmermans, and James B. Wiley. "Experimental Analysis of Choice." *Marketing Letters* 5, no. 4 (1994), pp. 3561–3567.

Dobber, David, and Ian G. Horgan. "A Comparison of Techniques Used and Journals Taken by Marketing Researchers in Britain and the USA." *Service Industries Journal* 8, no. 3 (1988), pp. 277–285.

Gardner, David M., and Russell W. Belk. *A Basic Bibliography on Experimental Design in Marketing.* Chicago: American Marketing Association, 1980.

Gaul, Wolfgang, and Christian Homburg. "The Use of Data Analysis Techniques by German Market Research Agencies." *Journal of Business Research* 17 (1988), pp. 67–79.

Naumann, Earl, Donald W. Jackson, and William G. Wolfe. "Examining the Practices of United States and Japanese Market Research Firms." *California Management Review* 36, no. 4 (1994), pp. 49–69.

Kuhfeld, Warren F., Randall D. Tobias, and Mark Garratt. "Efficient Experimental Design with Marketing Research Applications." *Journal of Marketing Research* XXXI, no. 4 (1994), pp. 545–557.

Steckel, Joel H., Wayne S. DeSarbo, and Vijay Mahajan. "On the Creation of Acceptable Conjoint Analysis Experimental Designs." *Decision Sciences* 22 (1991), pp. 435–442.

Index

About the Author

GORDON L. PATZER is Dean of the School of Business Administration, California State University, Stanislaus. In addition to teaching, Dr. Patzer is also a consultant to organizations around the world. He was previously the owner of two small businesses and earlier held marketing positions with CBS Television Network and the international advertising firm, Saatchi & Saatchi. Dr. Patzer is the author of *The Physical Attractiveness Phenomena* (1985) and *Using Secondary Data in Marketing Research: United States and Worldwide* (Quorum, 1995).