

TROUBLESHOOTING AUDIENCE RESEARCH

DETECTING PROBLEMS IN DESIGN, EXECUTION AND INTERPRETATION

Walter McDowell

National Association of
NAB
BROADCASTERS



TROUBLESHOOTING AUDIENCE RESEARCH

**DETECTING PROBLEMS IN
DESIGN, EXECUTION AND INTERPRETATION**

Walter McDowell, Ph.D.

Southern Illinois University



Troubleshooting Audience Research

TROUBLESHOOTING AUDIENCE RESEARCH: Detecting Problems in Design, Execution and Interpretation

©Copyright 2000

The National Association of Broadcasters
1771 N Street, NW
Washington DC 20036-2891

All rights reserved. No part of this publication may be reproduced or distributed in any form by any means, or stored in a database or retrieval system, without the express written permission from the National Association of Broadcasters. Requests for such permission should be addressed to the publisher.

ISBN 0-89324-289-6

*There are three kinds of lies:
damned lies, and statistics.*

Mark Twain

Troubleshooting Audience Research

Table of Contents

Chapter 1 Introduction	9
Chapter 2 Research as a Chain of Events	11
· Our Case Study	
Chapter 3 Troubleshooting the Design	12
· What Do We Want to Know? - Transforming Business Problems into Research Questions	13
SWOT Exercise	
Using Research to Enhance SWOT Exercises	
Using Focus Groups	
The Attitude vs. Behavior Dilemma	
· Who's in Charge? - Assigning Responsibility and Authority	21
Insist on Unconditional Honesty	
Selecting an Outside Research Firm	
Research vs. Consulting	
· Choosing the Right Research Method	24
Telephone	
Mail	
Face to Face (in-person)	
Group (auditorium)	
Internet	
Electronic Meters	
· Creating the Right Questionnaire	28
Methodology Considerations	
Vocabulary & Sentence Structure	
Memory Issues	
Attitude vs. Behavior Questions	
Vagueness and Ambiguity	
Order Effect	
Multiple-Choice Challenges	
To Rank or Not to Rank	
Measuring Strength of Response	
Multiple Dimensions Issues	
Anonymity and Confidentiality	

Troubleshooting Audience Research

·	Choosing the Right Sample Design	36
	Simple Random	
	Stratified	
	Systematic	
	Cluster	
	Sampling Error	
	Sample Size and Reliability	
	Sample Size vs. Population size	
	Weighting Samples	
·	Time, People and Money - Determining the Scope of the Study	41
	Cost Cutting	
	Setting Benchmarks	
·	The Devil is in the Details Part I - Troubleshooting a Research Proposal	43
	Criticizing Proposals	
	Great Expectations?	
Chapter 4	Troubleshooting the Execution	45
·	Taking a Test Drive - Pretests and Pilot Tests	45
	Using Focus Groups Again	
·	Would You Mind Answering a Few Questions? - Soliciting Cooperation	46
	"Just more junk mail"	
	"It's none of your damn business"	
	"What's in it for me?"	
	"Jeepers, this looks like it will take forever!"	
	"You want me to do this now?"	
	"I just couldn't figure out what they wanted"	
	Follow Through	
Chapter 5	Troubleshooting the Interpretation	53
·	The Devil is in the Details Part II - Troubleshooting a Final Report	53
	Cover Letter/Memo from the Researcher	
	Cover Page/Table of Contents/Introduction	
	Executive Summary	
	Background	
	Description of Method	
	Results	
	Conclusion and Recommendations	
	Oral Report	

- Reality or Smoke & Mirrors? - The Art and Science of Statistical Deception .. 57
 - Averages and Variances
 - Standard Deviation
 - Rankings
 - Percent Change
 - Optical Illusions
 - Correlation vs. Causation
 - Reporting Sampling Error

Chapter 6 Troubleshooting Nielsen and Arbitron 67

- Read the Fine Print 67
 - From Nielsen
 - From Arbitron
- Calculating Sampling Error 68
 - Nielsen Diaries
 - Nielsen Metered Markets
 - Arbitron Diaries
- Sample Composition 71
- Why Ratings Are Imperfect 72
- If You Still Want to File a Complaint 74

Concluding Remarks: Don't Give Up the Ship 75

Appendix A 77

Troubleshooting Audience Research

Chapter 1

Introduction

The adage that knowledge is power is no truer than in today's ever-changing and highly competitive world of telecommunications. Radio and television broadcasters face competition from not only their broadcast peers but from an array of alternative delivery systems. Whether advertiser-based or subscription-based, these electronic rivals all have a similar goal - to win the hearts and minds of audiences.

Complicating this battle for eyes and ears is the fact that as the number of program options has increased dramatically in the past decade, the amount of time people spend with conventional media in general has remained almost unchanged. More choice has not translated into more listening or viewing, which means that the only way for a broadcaster to survive is to maintain or increase share of market. The art and science of attracting audiences requires knowledge of their attitudes and behaviors. Acquiring this vital knowledge is no easy task. From innocent misunderstandings to intentional deceptions, the quest for genuine knowledge is a challenge for today's media decision-maker. Whether they're in the news, sales, programming or marketing areas, most media managers do not have a strong background in research methods and inferential statistics, yet are called upon to make important management decisions based on research *conducted by someone else*. A major problem today is that much of this research can be flawed in some way.

The consequences of poorly conducted research can be worse than the consequences of no research at all. Some media managers become the gullible victims of bad science, and consequently make bad business decisions. The old computer programming saying of "garbage in, garbage out" is appropriate also for audience research in that poorly composed research can lead to false or misleading results. But this abuse of quantitative research is nothing new. Over a century ago, Mark Twain quoted Disraeli in lamenting how numerical data could be exploited to support just about any misguided opinion when he wrote, "There are three kinds of lies: lies, damned lies and statistics."

The purpose of this book is to enable high-level broadcast decision-makers to recognize and overcome shortcomings in audience research. As the title implies, this book is not so much a "how to" text, but rather a "what to look out for" troubleshooting guide. While professional researchers may find this text of some value, the intended reader is the typical broadcast manager who is called upon to commission, supervise, and evaluate audience research conducted by other people, as well as all those who wish to understand this process better.

Troubleshooting Audience Research

The text begins with a basic overview of the research process and the chronological steps that must be administered properly for any research study to be meaningful. The later sections focus on specific trouble spots in the design, execution and interpretation of quantitative audience research. Whether a study originates from an in-house research department or an outside research firm, the basic principles of this troubleshooting guide can be applied to any type of mass media study. Both managers and researchers need to recognize that research alone does not produce solutions or strategies. Instead, *research produces knowledge* which, in turn, must be translated into action plans that will solve problems.

The chapters are formatted to include specific “*Warnings*” and specific questions to “*Ask the Researcher.*” By becoming attuned to the warning signs of flawed audience research and knowing what questions to pose to the designated researcher, a broadcast executive is increasing the odds of making the right research-based management decision.

A hypothetical case study is used throughout the text to provide some realistic examples of the type of research problems a broadcaster can face.

Chapter 2

RESEARCH AS A CHAIN OF EVENTS

Just as a chain is only as strong as its weakest link, so too are the results of a research study dependent on the proper implementation of several crucial steps, any of which can impair the accuracy of the study. These steps can be organized into the three basic areas of Design, Execution and Interpretation.

It is important to remember that these steps usually occur in a sequence and that the integrity of each link in our research chain is dependent in part on the integrity of all links that come before it. The cumulative effect of several “minor” infractions along the way can result in major problems with the ultimate conclusions of a study. Let’s look at an example.

The general manager of a radio station wants to learn more about listener attitudes towards “Big Al,” the morning DJ. He commissions a telephone survey by a somewhat disreputable local research company.

Problem #1

The *purpose of the study* lacks clarity, so the research company wrongly interprets the real needs of the station. This initial misunderstanding contaminates the entire study.

Problem #2

The sample of listeners chosen is not representative of the station’s market. Therefore, any conclusions derived from the study are unreliable.

Problem #3

The cooperation rate among participants in the study is only 50 percent due to busy signals, no answers or persons who refuse to participate. This adds an additional bias to an already imperfect sample.

Problem #4

Survey participants are confused over the wording of several questions. This causes answers to be inconsistent.

Problem #5

When analyzing the collected data, the research company uses the wrong statistical techniques, thereby alleging conclusions about the survey results that are simply not true.

Combining problems one through five, we can see that the entire research study is a recipe for a decision-making calamity. After squandering money on the futile study, the manager knows little more about the audience perceptions of his morning DJ. Even

Troubleshooting Audience Research

worse, the manager may be fooled into using the bogus data to make a poor management decision regarding the future of “Big AI.”

Taking the above scenario a few steps further, imagine the results of this study being forwarded to advertisers and professional media buyers. These research-conscious media executives use this “knowledge” to make important business decisions for their clients. The research epidemic spreads as the radio trade press reports the supposed positive results of the study, and consequently dozens of stations around the country change their morning shows to sound more like “Big AI.” A few weeks pass, and the Arbitron ratings indicate that listeners for all these copycat “Big AI” stations have dropped out of sight. Who is to blame? Why, it’s Arbitron, of course.

Warning: Beware of any weakness in the chain of events leading to a research conclusion. Any one step can damage the credibility of the whole study.

Conducting audience research poorly is easy to do. Conducting audience research properly, taking all the necessary precautions, is a bigger chore. “Quick and dirty” shortcuts will usually result in bad information and therefore bad management decisions. The bedrock of all troubleshooting guidelines is the recognition that every step in the research process plays a crucial role. Each link requires special diagnostic tools to evaluate its strength and its impact on the overall integrity of the research chain.

OUR CASE STUDY

For over a year, the household ratings for the early morning local newscast on a network affiliated TV station have been going down. Where the station was once the dominant market leader, it has now slipped to number #2 and the station’s group owner is not a happy camper. The General Manager has allocated substantial budget dollars for audience research. Before the News Director panics and overhauls the studio set and fires half the anchors, or the Promotion Manager buys a new jingle package, the GM insists on conducting meaningful research to diagnose *exactly* what is going wrong. She also finds the time to read *Troubleshooting Audience Research* published by the NAB.

The station has an in-house Research Director, an MBA with a good background in quantitative research methods and statistics. In addition to subscribing to the standard Nielsen sweep ratings, the market has just become a metered market, offering overnight household ratings. The station is also a client of a proprietary news consulting company called Frank Marvel & Associates.

Chapter 3

Troubleshooting the Design

WHAT DO WE WANT TO KNOW? - TRANSFORMING BUSINESS PROBLEMS INTO RESEARCH QUESTIONS

The first link in our chain concerns the initial design of the study. Depending on what questions need to be explored, research designs can vary. Our first goal should be to (a) identify a business problem and (b) transform this problem into a well-formed research question. Borrowing some guidelines from our colleagues in journalism, we can say that audience research provides answers to questions about:

Who?

What?

Where?

When?

Why?

How?

- Who listens or watches to your radio station after midnight?
- What do audiences claim are the most important elements of a sports segment?
- Where in your coverage area do the most affluent citizens live?
- When is the best time to schedule a new syndicated program?
- Why do audiences prefer your competitor's local news over yours?
- How can you create a more competitive program based on audience needs and desires?

From the research question evolves the research design elements. Professional researchers will often ask a client the following:

In one sentence, what is the question you want answered?

Surprisingly, many media managers cannot provide a quick, articulate response. The reason is usually a *poorly defined business problem* that is too vague and imprecise to convert readily into a simple straightforward research question. Researchers appreciate the old saying, "A problem well defined is half-solved." Let's return to our case study. The distraught News Director invites the station Research Director into his office. Let's eavesdrop on their conversation.

Researcher: "So what's up?"

Troubleshooting Audience Research

News Director: "We have a big problem, the ratings for the morning news are awful, we need some help."

Researcher: "And how can I help you folks?"

News Director: "Well, we need to figure out ways to attract more viewers."

Researcher: "And you want me to do what?"

News Director: "Help, of course!"

Researcher: "OK, what exactly are your research needs?"

News Director: "We need to find out more about the audience acceptance of the show."

Researcher: "Do you mean why people watch it or don't watch it?"

News Director: "Well, both... sort of"

Researcher: "Let me put it another way. In one sentence tell me what the question is that you want answered?"

News Director (after a very long pause): "Let me get back to you on that...meanwhile let's get cooking on some good research... Maybe some focus groups or maybe a survey at the Mall...Ah, maybe Nielsen can crunch some data for us. Let's get cracking!"

Warning: Poorly defined problems lead to poorly designed research. Solutions are only as good as the quality of the questions asked.

SWOT Exercise

The first link in our case study chain is weak because of a lack of *disciplined thinking*. The problems have yet to be properly defined. One way to discipline (not punish) management decision-makers is to walk them through a SWOT exercise. SWOT is an acronym for Strengths, Weaknesses, Opportunities and Threats and is often integrated into corporate strategic planning meetings. These concepts create a framework for analyzing all kinds of marketing situations. Notice that instead of dwelling entirely on negative issues, the format also encourages an examination of positive issues. Within this context a business "problem" can be interpreted to mean more of a *challenge*, such as how can a station better promote its strengths or how can the station management recognize an opportunity for audience growth.

The SWOT exercise can take several hours or days and should include all people directly involved with the issue. A top manager, such as the GM, should lead the exercise and keep everyone on track. Of course, the GM may be too close to the project and lack the

necessary objectivity to run such a meeting. Investing in an outside facilitator, such as a professional management consultant or trained academic, may be the most productive way to uncover core issues. The goal of the SWOT exercise is to eventually reach a consensus on the following items:

- What are the most important problems that need to be solved?
- Among these important problems, which require more information (research) in order to make intelligent management decisions?

Delaying a research study until the organization knows exactly what questions need to be asked can be a wise strategy over the long run. Rushed, poorly planned research can cause a company to squander time, manpower and money.

Warning: Do not proceed with a research study without a well-defined business problem and a resultant research question.

Some researchers prefer to use the term *objective* rather than *research question*, but the definitions are identical. For example, one could state that the research question for a study is “how many men ages 18 to 34 listen to a certain station?” Or we could say that the *objective* of the study is “to determine how many men ages 18 to 34 listen to a certain station.” The only real differences between the two phrases are in sentence structure and punctuation. In either case, the more specific the question (or the objective), the less likely the research project will get off track.

Let’s return to our case study. Instead of leaping prematurely into a quick research study, the general manager initiates a daylong SWOT analysis with her management team. She reminds the group that the overriding purpose is to learn as much as possible about the situation. In fact, some marketing organizations prefer to use the term *situation analysis* or *brand audit* to describe this learning task.

Using Research to Enhance SWOT Exercises

Problems and questions can have a reciprocal relationship where one contributes to the other in a steady progression towards learning more about a topic. Providing answers to questions can cause problems to be redefined. A redefined problem can then generate new questions. As an example, let’s drop in on our case study SWOT session.

Almost from the beginning of the session, the team realizes they do not have sufficient information on the ratings performance of the troubled newscast. The initial overall problem of a “loss of ratings” is too vague and there is an obvious need for more specialized audience data. The overall problem is redefined to the following:

- We need to know if certain segments within the total audience have suffered this ratings loss more than other segments.
- If certain segments have changed on our station, we need to know if there have been

Troubleshooting Audience Research

corresponding changes in the performance of our *competitors*.

Now the *research questions* can be formulated easily into the following:

- What have been the long-term ratings trends among key demographic groups?
- Similarly, are there significant correlations between our demographic trending and that of our market competitors?

Fortunately, this data is already available from standard Nielsen reports. Within a few minutes the Research Director retrieves this information coupled with some revealing graphs. The data disclose that indeed some demographic groups have experienced this rating loss more than others and that one competitor, in particular, has been the beneficiary of our loss. As the SWOT session continues, the dialogue generates new problems. Some members of the management team believe that another aspect of the overall problem is a poor lead-in program. This stimulates an intense discussion of “earning” viewer loyalties versus merely “inheriting” a big lead-in. Consequently more problems are defined.

- We need to know how much of our lead-in audience is retained versus how much is retained by our competitors.
- We need to know the degree of viewer loyalty for our program and our competitors’.

Again, using available data from Nielsen custom “audience flow” and “duplicated audience” analyses, the following research questions can be formulated.

- What is the lead-in ratings retention rate of key demographic groups for our station and those of our competitors?
- Over the course of a typical week, what proportion of key demographic groups watch a station’s morning newscast *exclusively*?
- Among the viewers of all morning newscasts, what proportion remain loyal to a station for *other newscasts* scheduled later in the day?

The SWOT exercise is put on hold for a couple of days while this information is obtained from Nielsen. When the data arrives, the group reconvenes and continues its discussions. Finally, there is a shared feeling of closure and consensus. The management team agrees on the core problems and the necessary research questions that will guide any upcoming research study. Their conclusions from the SWOT Analysis are:

1. The decline in household ratings appears to be most prevalent among men 18-49. This demographic has migrated over to a competitor. This is reflected in tune-in as well as lead-in audiences.
2. On the plus side, our station appears to have considerable strength in female demographics. These audience segments remain both large and loyal.
3. Without knowing why female audiences like the program, a strategy to regain male viewers could backfire if the same strategy alienates loyal female viewers.
4. News producers and promotion people never consciously targeted women. Instead, “it just happened.”

5. Because the competition appears not to be targeting any particular audience segment, the management team agrees that there is an opportunity to win the hearts and minds of some target audiences.

Now what? The SWOT exercise has enabled the management team to craft a more meaningful *problem statement*. They need to know how to (a) regain lost male viewers 18-49 from the competition while (b) reinforcing the loyalty of current female viewers. This statement stimulates the following *research questions*:

- Why have certain male viewers 18-49 abandoned the morning newscast?
- Why do female viewers in general prefer our newscast to others in the market?
- What strategies might work to regain male viewers without losing female loyalists?

The station Research Director can now concentrate his talent, time and resources on providing answers to these questions. A more thorough examination of SWOT analyses and other decision-making theories are beyond the scope of this book, but the author does want the reader to appreciate the importance of formulating problems precisely and providing appropriate research questions.

The above research questions were generated solely from a group of concerned managers as opposed to a group of ordinary citizens - people who watch television. Instead of dwelling exclusively on the management perspective, *focus groups* shift the viewpoint to an audience perspective, offering fresh insights about problems, questions and possible solutions.

Using Focus Groups

Although our final goal is to acquire *quantitative* (numerical) data about our station, researchers often turn to non-quantitative or *qualitative* research techniques to gain a better understanding of the issues at stake. To many people working in the industry, "qualitative" often means measurable demographic, lifestyle or psychographic data. However, to professional and academic researchers true qualitative research does not involve statistics of any kind. Instead, these techniques are intended to explore ideas and insights. For example, in-depth interviews and focus groups can be used to:

- Familiarize the researcher with core issues
- Clarify key concepts
- Set priorities
- Formulate objectives
- Anticipate problems in the design, execution and interpretation of a study

Well-conducted focus groups consisting of ordinary customers (or audiences) often reveal valuable insights that a manager-based SWOT exercise might not uncover. But while focus groups can be a highly worthwhile research tool in generating ideas, this technique is also the most misused form of audience research.

Warning: Beware of focus group results that dwell on statistical analysis rather than creative insights. Focus groups are about possibilities, not probabilities.

Originally designed to reveal new insights, many media executives have turned focus groups into a shortcut audience survey tool. Instead of going through the necessary steps of creating a large, randomly designed sample base, the researcher sidesteps these requirements by inviting a dozen people to sit around a table and chat about the station's programming. While focus groups can indeed help generate ideas, perspectives and insights, these small group gatherings should never be considered a reliable measurement of what the larger population thinks. Just because four out of eight people agree or disagree on something does not mean that 50 percent of the real population will respond in the same manner. Focus groups are not about headcounts or determining the proportion of people who feel a particular way. Warning flags should go up as soon as the decision-maker or researcher begins to use *statistics* to analyze the responses presented during a session.

A focus group is a *qualitative* research tool intended to explore unfamiliar territory. Rather than conducting individual interviews, the interactive group environment is intended to have a "snowballing" effect where one person's ideas will stimulate comments from the other group members. These guided conversations are recorded and later analyzed by the researcher. The formal results should be long on concepts and short on statistics.

Using our case study, several focus group sessions with ordinary viewers may reveal potential problems involving the morning newscast that a conventional management SWOT analysis might never disclose. Suppose the focus groups uncovered the notion that some men may not like the newscast because of the somewhat abrasive personality of one of the anchors. Is this audience opinion pervasive throughout the market? We don't know, but it is probably a factor that should be contained in a later *quantitative survey*.

Warning: Beware of a focus group moderator who fails to control adequately the group dynamics. Timid participants should be encouraged to speak up. Aggressive people need to be held in check.

Good focus groups require a professional moderator, someone who can guide the group dynamics, someone who can encourage participants to express their ideas and expose their feelings. Conversely, a competent moderator must control individuals who may try to dominate a conversation and intimidate others from speaking candidly.

The bottom line is that moderating focus groups is not for amateurs. The best way to troubleshoot a focus group is to be present during the session and to observe the behavior of the moderator first hand. All concerned managers should be invited to attend these group sessions so that they can experience all the "intangibles" that may not be conveyed properly in a formal report. While appropriate managers should be permitted to witness a

focus group session, they should be not be seen or heard by the group's participants.

Warning: Do not disclose to the participants the identity of the company that is sponsoring the focus group.

In addition to hiring a professional moderator, good focus group research requires anonymity in that the participants should not know the name of the company for whom the gathering is conducted. Studies have found that when the sponsoring organization is known, a "halo effect" can occur where the focus group participants become reluctant to say anything bad about the product or service under study. Furthermore, these studies have found that the cash compensation provided by the company reinforces their hesitancy to be overly critical.

Although some companies use focus groups exclusively to learn about their customers, the best way to capitalize on this type of research is to use these *qualitative* results to create better *quantitative* surveys. Sometimes a business simply does not know what to ask. There may be an obvious business problem, but getting at the root causes of the problem may be more difficult. In particular, the question of "why" something is the case can be challenging. Focus groups can assist in increasing the clarity of a proposed research study. In particular, focus groups can often help researchers unravel the often-complicated relationship between attitudes and behavior.

The Attitude vs. Behavior Dilemma

There are four essential types of Research questions.

1. **Statistical or demographic questions** - How old are you?
2. **Recall questions** - Can you recall the slogan for Channel 9 News?
3. **Behavior questions** - When there is important news happening in your community, to which station do you tune?
4. **Attitude questions** - Which station offers the best local news coverage?

A person's stated *attitude* can often be a good predictor of *behavior* - someone claims to like country music, and therefore one would expect that this person tunes to radio stations with a country format. On the other hand, a person's *behavior* can often be a good indicator of an *attitude* - a person who is observed to tune to country radio stations is presumed to appreciate country music. Market researchers often use a progression of consumer responses that begin with attitudinal measures and end with measures of purchase behavior, such as the following: *awareness, knowledge, intention, preference, motivation and behavior.*

Warning: Beware of pure attitudinal research. Audience attitudes are difficult to measure and therefore survey results can be disappointing in terms of predicting audience behavior.

Troubleshooting Audience Research

A problem facing audience researchers is that the relationship between attitudes and behavior is not as perfect as social scientists and market researchers would like. In general, people remain frustratingly unpredictable in that what they say does not always correspond to what they do. The broadcasting industry is filled with examples of attitude surveys that seem to contradict Nielsen and Arbitron estimates of actual tuning behavior. There can be many methodological reasons to explain these discrepancies and this book will examine several possibilities. However, the fact remains that the variable of human nature can be a major culprit in complicating the research process.

An example is peer pressure and the desire to present socially acceptable answers. This predisposition can cause people to give false answers on an attitude survey, especially when their responses are not anonymous or confidential. Other variables that can muddle the attitude/behavior equation can be external factors such as (a) the availability of a person to watch or listen; (b) the attractiveness of counter programming by competitors; and (c) the cost of acquiring the program content. For example, a potential TV viewer would prefer to watch a cable pay-per-view football game but he is not sure that he will be home to watch, and a competing channel is scheduled to air an equally attractive game for free. An *attitudinal* survey asking, "which game would you prefer to watch?" might provide misleading information as to actual tuning *behavior*.

Although the elements of many important consumer measures, such as awareness, preference, loyalty, and motivation, reside ultimately in the mind of the consumer, straight attitudinal surveys are notoriously unreliable. In many cases, researchers have found that behavioral questions can offer circumstantial evidence of underlying attitudes.

Let's use an example of head-to-head late evening newscasts, where each night of the week there is a different lead-in program. A researcher discovers that for one station, there is a strong correlation between the ratings of a particular newscast and the ratings of its lead-in; that is, the bigger the lead-in audience, the bigger the newscast's audience and *vice versa*. The program's performance seems dependent on its lead-in. Without asking people's attitudes about a news program, the researcher could deduce that this erratic viewer behavior reveals circumstantial evidence of *poor loyalty*.

On a more sophisticated level, custom studies provided by the ratings services can track audience behavior according to several factors such as "duplication," "shared listening" and "turnover ratio." Each of these behavioral measures can provide indirect attitudinal information about what people think as well as how they behave. A radio station that shares 70 percent of its audience with two other stations probably has a perception problem where listeners perceive all three stations as *equally satisfying*, and therefore not deserving of any real loyalty.

Going back to our case study, the in-house research director is aware of this attitude/behavior issue and recommends both types of measures. For example, Nielsen data could reveal patterns of behavior among male and female audiences that might shed light on their preferences. Additionally, a pure attitudinal questionnaire could delve into *why* people prefer a certain newscast. But before the details of a study begin to take form, everyone involved in the project must understand **who is in charge**.

WHO'S IN CHARGE? - ASSIGNING RESPONSIBILITY AND AUTHORITY

Determining who will actually conduct a research study depends on (a) the complexity of the proposed study and (b) the competency of an in-house research department. While large market stations may have the resources to properly design, execute and interpret a research project, many small market broadcasters must hire external research companies. A broadcaster may be tempted to try some in-house, low budget, do-it-yourself research, but the consequences could be disastrous. Remember that the words *cheap* and *inexpensive* have distinctly different meanings.

Insist on Unconditional Honesty

The working relationship between the management decision-maker and the researcher is crucial to the success of a project. Whether one is assigning the responsibility to an in-house research director or hiring the services of an outside vendor, there are some key ground rules that must be maintained.

Warning: Those responsible for conducting the research study may be intimidated by office politics and perceived hidden agendas. This situation can result in a lack of complete honesty in defining real problems and reporting results candidly.

Audience research is an exercise in science, not office politics. Therefore, the all-too-human frailties of sensitive egos and stubborn pride must be held in check. More specifically, conversations between the decision-maker and the designated researcher should be open, candid and deal with legitimate concerns. Whether addressing issues in news, programming, or marketing, department managers often become overly defensive and paranoid about the true intentions of a study. This is a challenge for top management to set the tone of cooperation and the need for good audience information. Research should be perceived as an impartial tool that will enable managers to make more enlightened decisions. If there is a real or imagined ulterior motive or a hidden agenda to a proposed study, top management needs to step in immediately and reconcile any problems.

In many corporate structures, the in-house researcher does not report directly to the highest authority within the organization (e.g. the general manager or a group president). Consequently, fear of some kind of reprisal from "upper management" can taint the actions of the researcher. Here again, top management needs to intervene and assure the researcher that he or she can work in an environment of honest inquiry without jeopardizing job security or future promotions. Earlier we stated that the first step in designing a research study is generating a well defined *business problem*, but the researcher's use of the word *problem* may disturb some people and result in a kind of cross examination - "who says there's a *problem* in my area?" The researcher, after analyzing the data, may be a bearer of bad news and be forced to endure another type of cross-examination - "So you think *you* know how to run a news department?"

Troubleshooting Audience Research

Similar problems may exist when hiring an outside research company. Because the station or broadcast group is compensating the researcher, there can be insinuations about the “expected” results. Most research companies want to cultivate repeat business and as a result, may be reluctant to disagree with a manager or to offer critical conclusions from the collected data. Again, top management needs to make sure that the integrity of the research is not compromised. In audience research, honesty is indeed the best policy.

Although the decision-maker may not be involved personally in the “nuts and bolts” of generating research reports, this person must demonstrate real leadership in protecting the soundness of the process.

Ask the Researcher: Do you feel any pressure to alter the proper way to conduct this study? Are you reluctant to be totally honest in presenting the conclusions of your research?

If the answer to either of the above questions is yes, the decision-maker must rectify the situation immediately. Claiming a lack of expertise in research methods is no excuse for avoiding this issue. The decision-maker is obligated to create a *politically neutral* atmosphere where the researcher is free to pursue the truth.

Suppose in our ongoing case study that there are rumors flying that the weather anchor is the primary cause for the loss in male viewers and that he will probably get fired in the next few weeks. The nervous anchor stops by the Research Director’s office for a little chat to remind the researcher that his anchoring job is “on the line” with this upcoming study. Later, the News Director, who hired the anchor, stops by and makes some obtuse comment about how colleagues on the same team “shouldn’t go around back-stabbing co-workers.” These comments may influence how the researcher conducts the study or reports the findings.

For our study, let us presume that the general manager is made aware of the rumor mill and immediately calls a meeting of all department heads. Her agenda is simple. One, there have been no decisions made regarding the future of any anchors. Two, there will be no interference in the proper planning, execution and interpretation of the upcoming study. Three, she reminds everyone that the station Research Director and any hired outside research firms will report directly to her. Four, she stresses that this research is intended to help the management team, not threaten it.

Selecting an Outside Research Firm

Unfortunately, audience research is not a licensed profession, so anybody can claim to be a “researcher” and the industry has its share of charlatans. Troubleshooting a research company that is soliciting your business requires time and some common sense.

Warning: Beware of any research company that appears reluctant to provide detailed client references or examples of prior work.

First, ask for references and actually talk to them. Too often people put references on resumes and proposals with the expectation that the reader will not follow through and actually contact these references in person. Research firms often present a laundry list of impressive client names but neglect to provide phone numbers or E-mail addresses. The decision-maker should request contacts and make a few calls. If the research company truly wants your business, extending this courtesy should not be a problem. A second recommendation is to request examples of prior studies done by the company.

Often a station or other media organization has a professional researcher on staff but still elects to hire an outside firm because the project is too big to handle in-house. In these situations, it is recommended that the staff researcher remain “in the loop” as an advisor for the management decision-maker and as a technical liaison to the research company. It is possible that the final project evolves into a partnership between the staff person and the external vendor where each is given certain responsibilities.

Research vs. Consulting

Many media organizations hire consulting firms to conduct audience research and make specific recommendations based on that research. The boundaries between research and consulting can become blurred when managers cannot differentiate between results and recommendations. Later in this book there will be a discussion of what a decision-maker should expect in a final report from a researcher. One vital criterion is a clear distinction between (a) the statistical results and (b) any recommendations derived from an interpretation of that data.

Let's take an example from our case study. Suppose the statistical results reveal that many male viewers are disappointed in the lack of sports and business news in the morning program. A recommendation in the final report states that the program add a fourth member of the anchor team who will focus on these two elements. After reading the report, the News Director is horrified at the thought and lashes out at the consulting firm for generating research of “dubious value.” However, the General Manager in our story and the station researcher take an impartial second look at the data analysis and conclude that there is indeed a genuine problem, but that the consultant's recommendation is only one of several options the station can take.

Warning: While consultants may be involved in the design, execution and interpretation of a study, these links in the research chain should be evaluated *independently*.

Consulting firms by definition give advice based on research data, but it is important to remember that the research methods and interpretation of results are separate links in the

Troubleshooting Audience Research

research chain. When troubleshooting audience research, the decision-maker must appreciate this separation and not succumb to overly broad criticisms.

So far, we have addressed the following:

(a) defining business problems; (b) transforming these problems into research questions; and (c) assigning responsibilities. While these elements of a study demand the active participation of the decision-maker, the upcoming elements require more work from the designated researcher. However, the decision-maker still must oversee and approve the entire proposal and therefore must maintain a critical eye on the remaining procedures.

CHOOSING THE RIGHT RESEARCH METHOD

There are several accepted survey methods for reaching people and each has certain advantages and liabilities. Among the most common methods used in broadcast research are:

- Telephone
- Mail
- Face to face
- Group (auditorium)
- Internet
- Electronic meters

Telephone Surveys

Surveys administered via the telephone have the distinct advantage of immediacy. Hundreds of calls can be made within a few hours and the data can be collected and analyzed right away. But there are also some trade-offs when using the telephone method.

Warnings for Telephone Surveys

Beware of the following problems:

- In some major cities over half the population has an *unlisted* phone number.
- Many participants will not tolerate long interviews.
- The timing of calls may be inconvenient, thus reducing cooperation rates.
- Younger people tend to be less cooperative than older people, thus introducing a possible bias.
- Highly visual topics are difficult to present orally over the phone.
- Complex multiple-choice questions are difficult to administer.

Even with the above-mentioned liabilities, telephone studies offer some unique assets. For example, the interviewers can offer personalized assistance if the survey participant is confused about the instructions or the meaning of certain questions. Because the calls are made usually from one central location, the interviewers can be monitored carefully, assuring quality control. In general, telephone surveys generate higher cooperation rates than other methods. Rather than pulling phone numbers at random from the phone book, a broadcaster can introduce computerized random digit dialing systems or hire the services of a professional market research firm. Broadcasters often use these research firms to do telephone coincidental surveys to supplement conventional ratings data provided by diaries or electronic meters.

Mail Surveys

Mailed surveys are often a viable alternative to telephone studies because there are no hourly labor costs for interviewers or expensive long distance phone charges. But mail surveys generate their own unique expenses in the form of printing and postage. In addition to budget considerations, there are several methodological problems.

Warnings for Mail Surveys

Beware of the following problems:

- Mailed surveys generate poor cooperation rates. They are often perceived as "junk mail"
- The correct addresses of highly mobile people, such as college students, are seldom contained in databases.
- There can be a long lag time from when the survey is received to when it is mailed back to the researcher.
- There is no opportunity for personal interaction with the participant to clarify questions.
- The researcher has no way to know who actually filled out the questionnaire.

On the positive side, mail surveys have some distinct advantages over telephone studies. First, mailed questionnaires can be fairly lengthy and complex. The participant can complete the task at his or her convenience and pace. Additionally, this convenience encourages more thoughtful responses where the participant has more time to deliberate. Nielsen and Arbitron use mailed diaries during their "sweep weeks." Households and individuals are solicited via telephone and then sent weekly diaries. A later section will focus on problems involved with diary keeping.

Face to Face Surveys

Face to face or "in-person" surveys are the most personal way to deal with participants and can range from simple shopping mall intercepts to long interviews inside a person's

Troubleshooting Audience Research

home. Of course, in exchange for this in-depth information there is a downside. Face to face surveys are inefficient in terms of time, people and logistics.

Warnings for Face to Face Surveys

Beware of the following problems:

- Interviews are usually time-consuming.
- Labor expenses can be high depending on how many interviewers need to be trained, transported and housed.
- Logistical problems can arise when surveys must be administered in faraway locations.
- Many public locations, such as shopping malls, forbid "solicitations" of any kind.

Although the above human and financial costs can be substantial, in-person interviews offer some special advantages. When encountering a researcher face to face, people are less likely to refuse participation and therefore cooperation rates are often very high. Additionally, visual aids, such as photographs and videotape, can be integrated easily into the survey format. Face to Face surveys also allow for more qualitative follow-up questions. Many television news-consulting firms use this technique where a trained interviewer will invest 45 minutes to an hour with one participant or family.

Group Surveys

Group or auditorium surveys are similar to face to face methods, but instead of dealing with one participant at a time, the researcher administers the survey in a group setting. What is being described here is not a focus group of eight or ten people, but a more structured environment consisting of dozens or hundreds of people. The efficiencies of a group-administered survey can also generate problems.

Warnings for Group Surveys

Beware of the following problems:

- True anonymity is difficult to achieve. Although the questionnaire may not require any identification, there can still be social pressure on participants.
- Invitations must be sent several weeks in advance to enhance cooperation.
- The inconvenience of having to travel to a specific location on specific day and time may lower cooperation rates.

Because of the size of the event, individual interviews are impossible but there can be some opportunities to clarify questions and perhaps have a limited discussion. Radio music consultants often use this format to evaluate specific songs and on air personalities using various audiovisual devices.

Internet Surveys

Online surveys using Internet web sites or e-mail can be very cost effective. Postage, long distance phone fees, and labor expenses can be reduced or eliminated. Furthermore, demographically targeted e-mail addresses can be obtained from private vendors. But as with other methods of reaching people, the Internet is far from perfect.

Warnings for Internet Surveys

- Unlike telephones or postal boxes, many people do not have access to the Internet.
- Cooperation rates for many studies have been low. People perceive inquiries as junk mail or "spam."
- There can be a long lag time from "received" to "reply."

Many radio and TV stations maintain their own databases of people who contact the station via the Internet. Interactive "listserves" and "chat rooms" also can be used to solicit information.

Electronic Metered Surveys

Meters are used primarily by Nielsen Media Research to record overnight TV ratings in addition to the periodic diary-based "sweep" reports. Approximately 50 of the 210 DMA markets in the U.S are now metered. The primary advantage of these patented "passive" electronic devices is that the researcher does not have to rely on the often-suspect memories of participants. The meter automatically records the channel that is on. Of course, the mere fact that a TV set that is "on" does not necessarily mean it is being watched.

Warnings for Metered Surveys

Beware of the following problems:

- Participating families may leave the TV set on all day, whether or not people are actually watching.
- Meters are intended to measure behavior only.
- No attitudinal information is provided as to how much the program is liked, etc.

Troubleshooting Audience Research

In an effort to combat the first problem, Nielsen's local market service gradually is replacing its passive meters with the more sophisticated *Peplemeters* that Nielsen currently uses for national ratings. These upgraded devices require participants to "log in" when they are watching a program, thereby providing much needed demographic data that the present passive meters do not provide. More information concerning Nielsen meters is available in chapter six.

CREATING THE RIGHT QUESTIONNAIRE

Bad questions generate bad answers. How a question is worded makes a difference, but so do many other factors such as how the question is introduced, what questions come before and after it, and much more. Earlier, we discussed how a lack of clarity in defining the objectives of a research study could lead to poor research questions. Now we tighten our focus and examine the actual questions as they would appear on a survey.

Questionnaire design is a science unto itself and should not be shrugged off as something unimportant. Because ordinary people will be dealing with the questionnaire, there is no reason why a broadcast manager can't detect a potential problem that the research "expert" overlooked. *Never underestimate the power of common sense in audience research.* When designing a questionnaire, anybody on your staff can be a worthwhile troubleshooter. The following pages offer some common warning signs.

Methodology Considerations

Earlier we discussed the advantages and disadvantages of different research methods in reaching people.

Warning: The *methodology* chosen for a study can influence how a questionnaire should be constructed. What works as a mailer may not work for a telephone survey, etc.

The choice of methodology can influence questionnaire design. For example, with telephone surveys, the participant is "blind" to anything that is difficult to remember or visualize. Conversely, with mail surveys, the participant does have the convenience of a visual document, but no one is on the line to clarify immediately any problems or questions.

Vocabulary and Sentence Structure

The phrasing of statements and questions in a survey can influence the nature of the responses they generate

Warning: The *vocabulary and sentence structure* may cause a person to provide different answers to a question. Beware of "loaded" sentences that may imply a preferred answer.

Resolving these types of problems doesn't require a master's degree in social science research methods, just some objectivity and common sense. Suppose you encountered the following question:

Studies have found that people possessing superior IQ scores tend to watch evening newscasts at least 4 times a week. Please indicate in the space provided how many newscasts you watch per week: _____

Does this question encourage a person to answer truthfully or is it "loaded" with a bias? Although the above is an extreme example, there can be more subtle situations where the researcher unintentionally introduces a factor within the wording of a question that may cause a survey participant to be less than forthright in his or her answer. The goal is to have a question perceived by the reader as completely *neutral* with no hint as to the "right" answer or the personal agenda of the researcher.

Sometimes a single word can alter the response to a question. One study that addressed the allocation of tax dollars asked people to indicate categories where they wished such revenue would be spent. One random group received a questionnaire that used the word "welfare" as one of the categories. Another similar random group received an identical questionnaire, except the word "welfare" was replaced by the phrase "assistance for the poor." The statistical results indicated that the word "welfare" generated far less tax dollar support than "assistance for the poor." Obviously, "welfare" had a negative connotation that influenced people's responses.

Memory Issues

The human mind can play tricks on itself. Psychologists know that people can provide incorrect answers to all kinds of questions dealing with time. It must be emphasized that we are not talking about deliberate lies, but honest recollections that are simply wrong. The most dangerous questions deal with *long-term memory*, where a person must recall something that occurred months or years prior.

Warning: People have flawed memories. Beware of questions that demand too much recall ability.

The ratings services learned long ago that one week is about the maximum length of time people can recall accurately their media activity. To ask people how many years they have watched Channel 9 news regularly is far more troubling than asking if they watched last night. Studies have found that most people suffer from "telescoping," where events are recalled to have occurred *more recently* than they really did occur. A solution for this recall problem is to have the question refer to specific dates or years. For example, instead of asking how many years ago a person graduated from college, ask precisely what year they graduated.

Another area of memory research that should be examined is the difference between *recall* and *recognition*. Questions involving pure recall are more challenging and the

Troubleshooting Audience Research

participant may not be able to offer an adequate answer. Recognition questions, however, provide some help that may “nudge” the participant’s memory. For instance, instead of requiring that participants recall all their favorite music performers, have them recognize names presented by the researcher on a comprehensive list of performers.

Attitude vs. Behavior Questions

In an earlier section we discussed briefly the relationship of attitudes and behavior and concluded that some attitudinal measures, such as preference and intention, often fail to be observed in actual market behavior.

Warning: Beware of research studies that dwell too much on highly abstract attitudinal questions.

Ignoring all attitude questions would be a mistake. Instead, the researcher can ground these more psychological questions in some type of *measurable behavior*. For example, instead of asking, “What do you consider to be important news?” ask, “When you choose to watch a local newscast, what type of news stories are you looking for?” The second question has a behavioral component (choosing to watch) that may offer more reliable results. Although many research questions, such as those found in our case study, begin with *why*, the measures used to answer the question may focus more on *how* audiences behave. For example, if our morning newscast introduces a more comprehensive sports segment and male demographic ratings go up, we can deduce that this *behavior* is an indirect measure of a *change in attitude* towards the program.

Vagueness and Ambiguity

Vague questions can be just as harmful as biased questions in that they create *confusion*.

Warning: Avoid vague questions that can be interpreted several ways by a survey participant.

Using our case study, suppose a participant is asked whether a station produces a “good morning newscast.” What do we mean by “good”?.... Compared to what? Would a participant interpret “good” to mean adequate or exceptional? A troublesome area in many surveys is the use of terms such as “often,” “seldom” or “occasionally” to denote the frequency of something. One person’s idea of “often” may not be the same as another person’s definition. It is sometimes better to present a range of *mathematical* choices, such as “1 to 5 times per week.”

Order Effect

There are dozens of studies that have explored this problem of question placement. The theory behind order effect asserts that some survey questions “set the stage” for

answering later questions. In other words, they make some answers more predictable than other answers. Psychologists have found that people tend to be consistent in their responses to most questions in life, so that once a pattern of answers has been established, they will continue to respond according to the pattern. Other researchers use the term “priming” to describe this phenomenon.

Warning: The order in which questions are presented may influence the way a survey participant thinks about an answer. Beware of questions that seem to “set the stage” for answering the next question.

Below is an example of three questions that suffer from order effects.

1. A recent national survey found the following items to be the most important issues facing America today: #1 crime, #2 racism and #3 unemployment. What issues do you believe to be the most important? _____
2. If you were news director of a radio or TV station, what types of stories would you emphasize on your regular newscasts? _____
3. Last night, station WXYZ aired a story on urban air pollution. On a scale of 1 to 5, how would you rate the importance of such a story? _____

The three questions interact and each contaminates the next question in line. Notice how question one “primes” question two. Imagine if each question was asked in complete isolation of the other two questions. Would the answers be the same as they are when presented in this pattern?

Another similar problem is *response sets*, where a pattern of questions encourages a set of *identical answers* from a respondent. Instead of investing some serious thought into each question, the person is tempted to take a “shortcut” and give the same simple response for everything. This patterned response is most commonly found when all of the questions deal with the same topic. For example, suppose the researcher administered several “agree/disagree” attitude questions dealing with the weather segment of our case study newscast.

1. The weather segment is the most important part of the newscast.

Strongly agree	Agree	Disagree	Strongly disagree
----------------	-------	----------	-------------------
2. Frank Freeze is my favorite TV weather personality in the area.

Strongly agree	Agree	Disagree	Strongly disagree
----------------	-------	----------	-------------------
3. I wish the weather segment came first in the newscast.

Strongly agree	Agree	Disagree	Strongly disagree
----------------	-------	----------	-------------------

Troubleshooting Audience Research

Suppose this survey continued with several more questions worded in the same manner, where each statement was *favorable*. Early on, a participant might become aware of this pattern or structure and *accelerate* through the remainder of the survey questions, giving the same response. A solution to this problem is to *break the pattern* by alternating favorable and unfavorable questions. For instance, we could rework question two in the following way.

Frank Freeze is my least favorite TV weather personality.

Strongly agree

Agree

Disagree

Strongly disagree

This alternating format will cause the participant to concentrate more and provide more thoughtful answers.

Multiple Choice Challenges

When dealing with a multiple-choice format, professional researchers insist that the proposed answers be “exhaustive” and “mutually exclusive.” In ordinary language, what they mean is that all possible answers must be made available and that these answers should not overlap in their definitions. Otherwise, the person taking the survey will become frustrated and offer an inaccurate response.

Warning: In a multiple-choice format, the *number and types of answers available* may force a participant to give an inaccurate response. Beware of too few choices and choices that overlap.

For instance, suppose a questionnaire had the following multiple-choice question:

Which of the following elements of a newscast is most important?

- (a) News about my community
- (b) Live on the scene reports
- (c) Sports scores & highlights

So what is wrong with this question? First, there may be another important element of a newscast that is missing from the list, such as weather - so the choices are not exhaustive. Secondly, choices (a) and (b) tend to overlap in that community news could also be presented live on the scene - so the choices are not mutually exclusive. The survey participant is forced into making an uncomfortable decision.

A corollary to the above warning of poor choices is the issue of “don’t know.” Researchers often presume that everybody has well-formed opinions or attitudes, when in reality, a person may be ignorant or unsure about the topic. Within a multiple-choice format, there can be circumstances when a “don’t know” or “no opinion” option should be available. By not offering this escape answer, the researcher is coercing the participant

to provide a response that may not be genuine. Several fascinating studies have found that many people do not want to admit that they “don’t know,” so they provide answers anyway! This was discovered by creating questions about totally fictitious people and news events (“How do you feel about Senator Fudd’s proposal to invade Iceland?”).

To Rank or Not to Rank

Rankings are a popular way to organize research data, particularly if an organization wants to brag about something. Being “Number One” is an effective marketing tactic. From a pure research perspective, ranking questions are simple to administer but there are a few precautions.

Warning: Ranking type questions are only as valuable as the list of items to be ranked. Beware of ranking questions that omit important factors.

First, remember that by definition, a ranking is a relative measure. That is, rankings compare one item against all items listed. Instead of asking about the importance of each item by itself, the researcher is asking about the importance of an item in relation to all the other items presented. As with multiple-choice questions, if the list of items to be ranked is not exhaustive and mutually exclusive, there can be confusion. Furthermore, if the participant regards two or more items with the same value, there is no way to resolve a tie.

Measuring Strength of Response

A survey that includes attitudinal questions should allow participants to express *how strongly they feel* about their responses.

Warning: The strength of a response can vary dramatically among participants. Beware of overly simplistic “yes/no” survey questions.

Instead of an imprecise “yes or no” response mechanism, the researcher should provide a strength of response scale, sometimes referred to as a Likert Scale.

Below are two examples.

Example A

Do you believe that news reporters should have access to the health records of convicted criminals? Yes _____ No _____

Example B

On a scale of 1 to 5, with 1 meaning totally disagree and 5 meaning totally agree, indicate your level of agreement with the following statement.

Troubleshooting Audience Research

News reporters should have access to the health records of convicted criminals

Totally disagree 1 2 3 4 5 Totally agree

A person may be “leaning” in one direction but not have a strong position on the issue. Example B offers more flexibility and therefore more precise insight.

Multiple Dimension Issues

Life is complicated, and although we often want to simplify our lives, there are times when we must cope with complexity.

Warning: Many important media concepts are multidimensional. Beware of overly simplistic questionnaires that attempt to deal with a complicated topic in only one or two questions.

For example, suppose a broadcaster wanted to learn about how audiences perceive the credibility of a news program. Below, questionnaire A represents the shortcut version. Questionnaire B represents a far more sensitive measure.

Questionnaire A

Please circle the number that represents how strongly you agree with the following statement (with 5 meaning you agree very much and 1 meaning you totally disagree).

Eyewitness News on Channel 9 is a credible newscast.

Agree 5 4 3 2 1 Disagree

Questionnaire B

Please circle the number that represents how strongly you agree with the following words concerning the credibility of Eyewitness News on Channel 9 (with 5 meaning you agree very much and 1 meaning you totally disagree).

Impartial	Agree	5	4	3	2	1	Disagree
Accurate	Agree	5	4	3	2	1	Disagree
Trustworthy	Agree	5	4	3	2	1	Disagree
Factual	Agree	5	4	3	2	1	Disagree
Sensational	Agree	5	4	3	2	1	Disagree
Biased	Agree	5	4	3	2	1	Disagree
Confusing	Agree	5	4	3	2	1	Disagree
Irrelevant	Agree	5	4	3	2	1	Disagree

Returning to our TV case study, which style of question would be best if the management team wanted to better understand how much viewers appreciate the morning weather anchor? We could use one simple question addressing the degree to which he is either “good” or “bad,” or we could introduce a more sophisticated structure, offering a range of dimensions coupled with strength of response scales.

Competent	Agree	5	4	3	2	1	Disagree
Authoritative	Agree	5	4	3	2	1	Disagree
Friendly	Agree	5	4	3	2	1	Disagree
Understandable	Agree	5	4	3	2	1	Disagree

Sometimes, researchers use the above numerical scale in reverse order, where 1 represents the highest score (similar to ranking something #1) and 5 represents the lowest (as in 5th place). But combining scales with rankings in the same questionnaire can lead to confusion. A respondent can misinterpret the instructions and provide an answer that is the exact opposite answer from what he or she intended!

Other structural issues surrounding the use of scales concern the *number of options* offered to the participant. Our examples in this book have used a five-option structure (5 4 3 2 1), but some researchers prefer a seven-option format (7 6 5 4 3 2 1) as the perfect number. There is also considerable debate over the use of *odd versus even* number of options to measure middle positions. For instance, by choosing the number 3 on a five-option scale, a respondent can register an exact middle or ambivalent position. However, if the same respondent encounters a 4-option or 6-option scale, he or she is forced to “lean” in one direction or the other.

Anonymity and Confidentiality

Depending on the nature of the questions asked, a participant might be tempted to give false answers. Highly sensitive topics such as sexual habits or illegal behavior will discourage honest responses unless there is a way to protect the identity of the participant.

Warning: A lack of anonymity or confidentiality may influence the credibility of a participant’s answers.

Anonymity means that the researcher cannot identify who said what in a survey. An example would be a “blind” questionnaire that does not request a person’s name or any other identification. Once the questionnaire is handed in, there is no way to trace its origins. Confidentiality means that the researcher can identify a participant’s responses but this information will not be released in any public reports. An example would be surveys conducted by Nielsen and Arbitron, where the research companies know who filled out a diary or is a member of a metered family but does not release this information to anyone outside the company.

Troubleshooting Audience Research

These and a score of other design factors can result in wrong conclusions about the target population. An antidote for many questionnaire problems is to “test drive” the survey with a small group of people who can provide candid feedback.

CHOOSING THE RIGHT SAMPLE DESIGN

Because it is impossible to survey everyone in the population, we are forced to acquire a representative sample of the desired population and infer conclusions from the sample back to the “parent” population. The study of *inferential statistics* is basically the study of samples and how accurately they represent a larger population. When dealing with sample-based audience research, a well-designed random sample is essential. Random means that every unit within the targeted population has an equal chance of being chosen. Random samples are often referred to as *probability samples*, implying that they are dependent on the laws of probability. The most insightful questions coupled with the most sophisticated statistical analyses are still worthless if the sample is biased. A *biased sample* is one where certain persons within the population have a greater likelihood of being selected than others are. Biased samples are often referred to as *convenience samples* meaning that they are generated with no way of knowing if those persons included in the sample are representative of the target population under study.

Warning: Convenience samples can be biased. Beware of researchers who do not appreciate the need for randomness in a proposed study.

There are several established ways to acquire a random sample and a management decision-maker should be familiar with the basic elements of each of these procedures.

- Simple random
- Stratified
- Systematic
- Cluster

Simple Random Sample

As the name denotes, a simple random sample is derived by the simplest means available, such as a blind drawing of names out of a hat or using a mathematical table or computer program to generate random numbers.

Stratified Samples

A stratified sample is again a probability sample, but is divided among several predetermined subpopulations or strata. Most strata are demographic categories such as age, gender, education and income. A stratified sample requires that a simple random sample be drawn independently from each stratum. Broadcasters interested in knowing more about the attitudes and behavior of various demographic groups would normally want such a sample design.

Systematic Samples

A systematic sample is a probability sample but introduces a systematic means of selecting participants rather than the more time-consuming pure random sample procedure. Instead of drawing a different random number for each person selected to be in a sample, the researcher uses only one random number and repeats it. For example, in a telephone survey format, the researcher will select every 10th name throughout the entire phone book. We can see that this systematic technique can be applied to a stratified sample, where each stratum is subjected to a systematic formula for randomness.

Cluster Samples (Multi-stage Cluster samples)

Big studies, such as those devised by Nielsen and Arbitron ratings services, make use of a staging technique where a target population is subdivided progressively into ever-smaller random units or clusters of people. In other words, the sampling process takes place in stages where the final stage determines those who will be included in the measured sample. Let's demonstrate a multi-stage cluster sample for a large city. Instead of attempting a random sampling of every single street in the city, the researcher selects randomly a handful of boroughs. From these boroughs, the researcher draws another random sample of city blocks. Specific streets within these blocks are then selected randomly to be in the actual survey. Sampling by clustering is sometimes called *area sampling*.

The above has been a highly simplified discussion of sampling methods. Professionally trained researchers should have a far greater understanding of these concepts. However, the decision-maker should be familiar with the jargon of sampling and the basic principles involved in the effort to acquire a representative sample.

Ask the Researcher: What type of sample design do you intend to use for this study? How are you assuring randomness?

Regardless of the sampling technique proposed for a study, there are some important principles that apply to all situations employing a random sample.

Sampling Error

Even when a broadcaster decides to do a legitimate survey using a large well-designed random sample, there is still no guarantee that this sample is a perfect replica of the target population. Random sampling is based on the laws of probability, meaning that over the long run certain happenings can be predicted within certain boundaries. Casino operators deal with games of chance, yet they can predict their weekly earnings with great accuracy. Of course there will always be margins of error, but the good news is that these margins can be predicted mathematically.

A calculated *margin of error* will vary according to two basic factors: (a) the size of the sample and (b) the variability of the answers provided. Professional researchers will use

Troubleshooting Audience Research

statistical terms, such as *standard error* and *confidence intervals*, to gauge the probable error contained in a piece of data. These predictions are based on the laws of probability, which assume perfect randomness. If a sample is biased (that is, not random), these calculations become meaningless.

It is vital that the decision-maker has at least a cursory understanding of sampling error and the impact it can have on management decisions. One important rule of thumb is that in general, small samples are less reliable than big samples. Additionally, small estimates are less reliable than big estimates. Combining these two factors can result in a real train wreck. When examining (a) small demographic results from (b) tiny subsamples (such as a radio station earning a 1 rating in women ages 12 to 24 during late night hours), the sampling error can be so huge that the manager would be foolish to base an important management decision on such unreliable data.

Without getting overly embroiled in statistical jargon, there are some sampling statistics that a media decision-maker may find handy when working with an in-house research director or a hired research firm. Remember, we are dealing with estimates, not perfect measures of a population. These sampling statistics are presented in Chapter 4, *Troubleshooting the Interpretation*.

Sample Size and Reliability

The author promised that this handbook would not contain a great deal of math, but there is one area that requires an appreciation of some high school mathematics. Remember square roots?

The most straightforward way to increase the reliability of a sample is to increase its size, but there is an unfortunate twist to this solution. Sample error and sample size do not have a nice one-to-one inverse relationship, where if one goes up, the other goes down by the same degree. Instead, standard error is a function of the *square root* of the sample size. At first, this may seem illogical, but nonetheless, it is true in the world of mathematical probability. Look what happens when we keep squaring the numbers below. We begin with 4 and make a small jump to 16, but the next progression of squared numbers really takes off.

$$4 \text{ squared} = 16$$

$$16 \text{ squared} = 256$$

$$256 \text{ squared} = 65,536$$

Etc.

Rather than a simple direct relationship, we have an *exponential* one, which causes some big headaches for survey researchers. For instance, doubling the size of the sample does not reduce the error by half. Instead, according to this square root formula, you must quadruple the size of the sample. This mathematical quirk has a profound impact on the economics of sample-based audience research.

Warning: Beware of mathematical traps. Remember, sampling error is a function of the square root of the sample size, meaning that the researcher soon reaches a point of diminishing returns.

Although it is always possible to improve the accuracy of a survey by increasing the size of the sample, you reach very quickly a *point of diminishing returns* where to make the slightest increase in reliability requires a huge increase in sample size. A simple example would be where we have a starting sample size of 10 people and we want to reduce our sampling error by half. This means we must quadruple our sample size to 40. Now suppose we want to again reduce our error by half. This goal requires us to again quadruple our sample size to 160. In other words, while our sampling error is *4 times smaller*, our sample size is *16 times bigger*. You can imagine the problems researchers face when dealing with samples sizes that are in the hundreds or thousands. Suppose our original sample consisted of 1000 people (instead of 10) and we wanted to reduce the potential sampling error by half. Is it worth the time, people and money to survey 4000 people? Because of the nature of squaring numbers, as the size of a sample increases, it has progressively less and less impact on reliability. This is why most national surveys use samples that are seldom larger than 5000. The additional cost factors in administering 6000 rather than 5000 interviews may be thousands of dollars, but the net gain in improving sampling error (reliability) is tiny. When confronted with the mathematical reality of sampling design, most companies do not want to invest in gigantic samples that yield half a percent increases in precision. Later we will see that sample size has become a burning - and misunderstood - issue with the ratings services and their subscribers. For markets that currently use several hundred sweep diaries or meters, a 10 or 20 percent increase in sample size will yield only a tiny fraction of improvement in sampling error.

Sample Size vs. Population Size

Another aspect of sampling that perplexes many people is the relationship of the size of the population versus the size of the sample. There is an intuitive sense that in order to maintain sample reliability big populations should require big samples, but this is a false assumption.

Warning: Beware of a researcher who claims that sampling error is related in any way to the size of the population under study.

Pretend that you are cooking up a batch of your famous chili and you want to sample periodically this concoction using a small spoon. Does the size of the spoon have anything to do with the size of your pot of chili? Of course not. As long as your spoonful provides a representative sample of the chili, the size of the pot is not relevant.

All things being equal, the size of the population under study has no direct impact on sampling error. The precision (sampling error) of a random sample remains the same, whether the parent population is 3,000 or 30,000. What does matter is the size of the

Troubleshooting Audience Research

sample compared to the variability of the responses. The exact formula and a more detailed discussion can be found in any introductory statistics text.

For radio and television markets, sample sizes are calculated according to the complexity of the market, not its size. That is, large markets require larger samples not necessarily because they are *big* but because they tend to have more media choices (stations, channels, etc.) and more diverse demographics than smaller markets. More choice and more diversity translate into greater variability in survey responses, which in turn threatens the reliability of the sample. If a large population is not demographically diverse and offers only a few media choices, a large sample is not necessary. Conversely, a relatively small population still requires a minimum sample size to control sampling error. Don't become trapped by the false notion that small populations need small samples.

Warning: Beware of a researcher who attempts to cure a biased sample by merely enlarging it.

In addition to its size, the *degree of randomness* of a sample is equally crucial. A big *biased* sample is no more representative of a population than a small *biased* sample. When generating samples, bigger isn't always better. All mathematical formulas for measuring sampling error presume that the researcher is dealing with a *perfect random sample*.

Weighting Samples

If the demographic composition for the target population is known, a researcher can evaluate the randomness of his or her sample by running mathematical comparisons. For instance, if the population under study is known to be composed of 45% males and 55% females, the in-tab sample should reveal the same proportion. If this is not the case, the researcher has two options. The first is to return to the field and recruit more of the desired demographic group. The second, and more common option, is to apply some *statistical weighting* to the current sample. If a sample appears to have too few women, the researcher applies a proportional weight to the female responses. That is, their responses are counted more heavily when analyzing data and presenting results.

Warning: Beware of samples that have excessive statistical weighting among demographic segments.

While this weighting procedure gives a sense of mathematical fairness to a study, there can be serious problems when interpreting the results. The researcher is assuming that this smaller, statistically enhanced segment of the sample is still reasonably representative of the original population, and that the few people who did participate in the study are still *typical*. The word "reasonably" was used in the prior sentence because we know that there is always a statistical margin of error associated with any sample-based study. Introducing demographic weighting, however, complicates things. Recall that standard error calculations are based on the assumption of a perfect random sample. Statistical

weighting is a mathematical device that is used to “cure” those samples that fall short of ideally reflecting a target population.

Are the collected responses also *representative* of those people who could not or would not cooperate? If the data obtained from this disappointingly small sample segment is biased, statistical weighting will only make things worse. For example, several studies have found that persons with strong opinions tend to volunteer for surveys more often than people who hold more neutral opinions on controversial issues. If, in a survey sample, the number of members of a particular demographic group is deficient, the researcher will give greater weight to the responses that have been turned in. The problem is that these collected responses *may already be biased* and now they are going to be exaggerated even more. A later section of this book looks at strategies to increase cooperation rates and avoid these potential biases.

If sampling weighting is used in a study, the calculation for standard error becomes more complex. Depending on the how deficient a particular demographic sample is and the amount of weight that is applied, sampling error can be enormous. While the overall reliability for all participants *combined* may be tolerable, the reliability of data generated from a breakout of narrow demographic subgroups may be unacceptable.

Ask the Researcher for full disclosure of any *sample weighting* and the appropriate *sampling error* for each demographic group.

TIME, PEOPLE AND MONEY - DETERMINING THE SCOPE OF THE STUDY

Good audience research requires an investment in time, people and money. As with most business decisions that involve financial and human costs, the decision-maker must look for a proper return on investment (ROI) - how big an investment is necessary to gain the desired return? Another way of looking at this issue is asking *what is the payoff?*

Cost Cutting

There is no limit as to how ambitious and expensive a research study can be. Overzealous research directors or greedy outside consultants can recommended giant studies that will generate more information than you ever wanted to know. A decision-maker faced with limited research budget dollars must be prepared to reduce the extravagant “bells and whistles” of a proposed study without violating its true purpose. This is yet another form of troubleshooting. Although we are not necessarily searching for flaws in the research design, we are looking for areas of needless expense - items that will not contribute significantly to the expected payoff. Here are some factors to consider.

The Experience and Expertise of the Researcher: Can the research objectives be achieved utilizing in-house personnel and facilities? For example, a well-trained station research director may be competent enough to design a telephone study and then recruit station staff members to place the calls from their desks.

Troubleshooting Audience Research

The Proposed Methodology: Depending on the circumstances of a study, certain methods may be more costly than others. For example, a survey that requires long distance telephone charges may be administered less expensively through the mail. As we discussed in a prior section, there will probably be trade-offs that may or may not hurt the objectives of the study. A shift from telephone to mail may save money but the timeliness of the results will be endangered. Also, long, in-depth questions used in a face to face survey may not translate well into a telephone questionnaire.

Ask the Researcher: Is there an alternative method that will deliver comparable results for less money?

Looking at our case study, we can see a tension between the in-house Research Director and the station's outside consultant, Frank Marvel Associates. The in-house person believes he has the ability and resources to do the study at half the proposed cost of using the outside firm. Frank Marvel insists that the job is too sophisticated for "local" people to execute properly.

The sample size: Requiring fewer people in a sample can result in less money spent on a study, but as we learned in a prior section, smaller samples generate less reliable results (higher standard error, wider confidence intervals). But remember too that there is not a direct one to one relationship between sample size and the degree of reliability. That darn square root business has a huge impact on increasing sample reliability. Using some simple algebraic formulas, a competent professional researcher should be able to determine precisely this statistical trade-off.

Ask the Researcher: How much reliability will be lost if we reduce the sample size by x percent?

Setting Benchmarks

Deciding whether a particular programming or marketing strategy has actually worked requires some kind of comparison. For instance, how do you know if a particular promotion campaign heightened awareness of a program or not? The only sensible way to evaluate change is to compare your current data with an established benchmark or baseline. In most cases this means *replicating the research study several times* over a period of months or years. In terms of positioning a program or station, repeated studies can reveal valuable information on competitive maneuvers over time.

When budgeting a research project, a wise manager should consider replicating the study in order to take full advantage of the results. The pay-off in knowledge gained from investing in two or more consecutive studies, rather than just one isolated project, can far outweigh any added costs. A troubleshooting guideline for someone commissioning a major research study is to make sure the proposed study can serve as a viable benchmark for additional studies.

The GM in our case study understands this issue and decides to initiate a benchmark study right away to see how the program is perceived today and a second study six months from now to evaluate any change. She also decides to “go in-house” for the study. The consultant is crushed, but that’s life, Frank.

THE DEVIL IS IN THE DETAILS PART I: TROUBLESHOOTING A FORMAL RESEARCH PROPOSAL

Warning: Beware of incomplete or confusing study proposals. Ask for missing items or clarification. Never presume anything.

Whether a study originates in-house from a research director or out-of-house from a professional research company, the decision-maker should expect the following items in a formal written proposal:

- A statement of the business problems and resulting *research questions* (objectives) of the study.
- A clarification of *responsibilities* and chain of command.
- A statement of *feasibility* from the designated researcher that assures the decision-maker that there are no qualms about the design or execution of the study.
- A specific *timeline* as to when certain components of the study will be completed.
- A thorough discussion of the *research methods* to be used, including the type of survey, sample design, and data collection method.
- A presentation of the actual *questionnaire* (including any instructions) that will be administered to the sample.
- A brief explanation of the *statistical analyses* that will be performed on the collected data. This is an area where the decision-maker may want to get a “second opinion” from someone who has a strong background in statistical analysis. The last section of this book addresses some common statistical tricks and traps encountered in audience research data.
- An *itemized budget* that includes not only financial costs but also demands on in-house people and facilities.

Criticizing Proposals

A proposal is just that - a proposal. Earlier we asserted that the research process is a chain of events in which each link must carry its own weight, but in reality, some links are more important than others. One of the most important links requires the sound judgment of the decision-maker - the *approval* of the research proposal. The stakes are too high for a manager to “rubber stamp” a proposal. Even when the proposal contains unfamiliar research jargon and esoteric comments, the decision-maker must have the courage to say, “I don’t understand,” and seek assistance.

Troubleshooting Audience Research

Great Expectations?

A final note is in order here. The decision-maker should not expect absolute perfection. Broadcasters and media buyers often forget that even in the best of circumstances, quantitative research results, such as those generated by Nielsen and Arbitron ratings measurements, are filled with unavoidable *sampling error*. All accredited audience research services recognize sampling error as a fact of life but all too often their clients do not. Keep in mind that certain fluctuations in the published ratings performance of a radio or TV program may be the result of mere random chance (“statistical bounce”) that has nothing to do with the genuine performance of the program or the research methods employed.

Chapter 4

Troubleshooting the Execution

TAKING A TEST DRIVE: PRETESTS AND PILOT TESTS

Just as you would never buy an automobile without a test drive, neither should a researcher administer a questionnaire to a large sample until it goes through a thorough “test drive.” Professional researchers refer to these preliminary actions as *pretests* or *pilot tests*.

Warning: Even professional researchers can miss a critical problem associated with a proposed questionnaire. Pretests or pilots tests are recommended highly to intercept potential problems.

Using Focus Groups Again

The goal here is to receive feedback from a small group of volunteers who complete the proposed questionnaire and then discuss any problems they encountered. This is an ideal venue for a focus group, where a moderator can guide the conversation to help fix a faulty questionnaire.

This may be the second time a focus group has been used to enhance the value of the research study. In our case study, during the problem definition stage, the General Manager commissioned some focus group sessions to supplement the information gleaned from the managers’ SWOT exercise. Now focus groups are brought in again, this time to fine-tune the questionnaire design. Of course, a test drive is not absolutely necessary, but imagine the consequences if a real “fatal flaw” was discovered in the questionnaire after it was administered to hundreds or thousands of people! Here are a few common sense guidelines for conducting a pretest.

1. Administer the questionnaire to the same type of person who will complete the approved version. Asking a group of Ph.D.s to answer questions intended for blue collar workers would likely not be productive.
2. Encourage people to speak up and be critical. Don’t allow a “halo effect” to hamper candid conversation.
3. Keep the feedback concentrated on the questionnaire itself, not the topic of the questionnaire. Unlike prior focus groups, this one has a very narrow focus - the viability of the questionnaire as a survey tool.
4. In addition to criticisms, ask the group to provide possible solutions to questionnaire problems.

Ask the Researcher: Have you pretested the questionnaire with an appropriate group of people? What changes have you made?

Postponing the administration of a major survey by a few days in order to double check the questionnaire design may be a very shrewd decision in the long run.

WOULD YOU MIND ANSWERING A FEW QUESTIONS? - SOLICITING COOPERATION

Many audience researchers agree that the number one problem facing survey research today is *poor cooperation*. For more than a decade, Americans have become ever more reluctant to participate in any kind of market research, including Nielsen and Arbitron ratings surveys. Instead of using the term cooperation rate, some researchers will refer to this measure as *refusal rate*, *response rate* or *nonresponse error*. Regardless of the preferred vocabulary, the problem is serious. While statistical weighting is introduced often to compensate for demographic groups that are underrepresented in an in-tab sample, there remains a nagging question: Are the responses from people who *do* cooperate truly representative of the people who *do not* cooperate?

If, indeed, the answer to the above question is no, you have introduced a *bias*. The “Catch 22” dilemma to all this is that it is impossible to know for sure (How can we know if these two groups are different when one of the groups refuses to provide the needed information!). Increasing the size of the sample will not solve the predicament. Fifty percent of 100 has the same biasing influence as 50 percent of 1000.

Ask the Researcher: What steps have you taken to maximize cooperation rates?

The best way to combat this type of sample bias is to find ways to *maximize cooperation rates*. Depending on the type of study, the researcher can use several strategies. Although the media decision-maker may not be responsible for the actual solicitation of sample participants, he or she should still “look over the shoulder” of the designated researcher and do some common sense troubleshooting.

“Just more junk mail” or “Why should I chat with this creepy looking guy?”

The very *first contact* with a potential survey participant should be positive and professional.

Warning: Beware of surveys that are perceived as *unprofessional*. First impressions matter!

For mail surveys, the physical appearance of the envelope and questionnaire should look businesslike, exuding respect for the person's time and consideration. Attention-grabbing graphics and "teaser" copy, common among most direct mail marketing pieces, may turn people away. A well-crafted cover letter from a top decision-maker in your company can help enhance the professionalism and seriousness of the study. Appendix A offers a sample letter and some guidelines.

In-person interviews, such as shopping mall intercepts, face the same problem of professionalism, only in these situations, the "packaging" includes the person representing the survey organization. The physical appearance and demeanor of an interviewer can make or break a solicitation. For example, most mature adults will be reluctant to invest much time talking to a disheveled teenager who starts a conversation by saying "Hey, man...you got some time to kill?"

Telephone interviewers have only a few seconds to convince the other person on the line that the survey is a legitimate enterprise and that the caller is not trying to sell time-share condos or magazines (studies have shown that telemarketing has damaged the reputation of professional market research). The interviewer must establish quickly a professional relationship and persuade the other person that his or her responses serve a valuable purpose.

A common strategy designed to enhance cooperation rates for mailed surveys is to first contact the designated family by telephone, alerting them to an upcoming mailer. Both Nielsen and Arbitron use this two-fold technique to recruit "sweep" participants. The firms also carry out several follow-up phone calls and postcard mailers to reinforce cooperation.

"It's none of your damn business"

Fortunately, most media surveys do not delve into highly sensitive matters, such as sexual habits or personal phobias, but there can be a few situations where the potential participant believes that your media organization is entering sacred ground. The obvious solution is to offer anonymity or at least confidentiality. A parallel strategy is to tell the participant how *important* the data will be for your project.

Warning: Beware of surveys that appear to invade a person's privacy without offering anonymity.

Perhaps your news department wants to conduct a survey about spouse abuse. By offering total anonymity and explaining how important the news report will be in revealing this social problem, a reluctant potential participant may become more open. Here again, the *professionalism* of the solicitation is crucial for success.

"What's in it for me?"

Good question: Why should someone invest time and energy in completing your questionnaire?

Warning: Beware of surveys that offer no tangible or intangible payoff. People want to be appreciated.

There must be some kind of positive payoff, even if the reward is psychological. Here are three different approaches, each of which has a possible downside.

First, *try flattery*. Many people love to express opinions on just about anything, especially if the receiver appears fascinated. Nielsen has exploited this with some success by referring to participating sweep families as "members" of an exclusive group of Americans who are shaping the future of television. As was mentioned in our section on focus groups, researchers need to be wary of the "halo effect" where the respondents are so flattered for being selected they become reluctant to say anything unkind about a survey topic.

Another approach is to *persuade the person that your study is relevant to their lives*, that someday, the results may be used to make their lives better - better newscasts, better features, better movie titles, better song titles, better web graphics etc. For highly controversial topics, people may be eager to express their opinions and the payoff is merely the opportunity to "go on record." A possible downside to this approach is that the researcher ends up dealing with people who have atypical or exaggerated opinions. For example, surveys dealing with the abortion issue often attract disproportionate numbers of people who have extreme views, while people who are neutral or ambivalent about the issue tend not to be as eager to participate in the sample. The result is a study that reveals no statistical "middle ground," when, in reality, the population is not as polarized as the sample-based data insinuate.

If enhancing a person's self-esteem doesn't do the trick, one can always revert to *incentives*, better known in some circles as bribes. Incentives can range from something as innocuous as a free slice of pizza to something more substantial, such as a chance to win tickets to a big concert or even cold cash. The down side to this strategy is the potential of introducing some kind of unintentional bias. *The key is to offer an incentive that has no obvious connection to the topic of the study*. Suppose a radio station conducted a survey about general music preferences and offered as an incentive a chance to win tickets to an upcoming jazz concert. Would the incentive introduce a sampling bias? Would someone who does not like jazz be less willing to cooperate in the study?

"Jeepers, this looks like it will take forever!"

This problem goes back to our section on questionnaire design. Studies have found that there are *time thresholds* when people will become progressively less cooperative.

Warning: Beware of surveys that appear time consuming. There is a direct connection between duration and cooperation.

Many people will refuse to cooperate if the survey will take more than seven or eight minutes of their time. As with the introduction of incentives, there can be an unintended sample bias when too many people refuse to participate due to the perceived duration of the survey. Perhaps the people who do agree to complete a long questionnaire have unusual opinions or behaviors that are not representative of the target population. Both Nielsen and Arbitron agree that older people tend to cooperate more often and are more conscientious than young people about filling out a complex weeklong diary. Likewise, people who have exceptional expertise or have intense feelings about a survey topic may be more easily persuaded to participate in a time-demanding survey. For example, women are far more willing than men to participate in long surveys dealing with relationships.

Rather than concocting ways to cajole or bribe people to complete a time-consuming survey, a better solution is to *sacrifice some of the less important questions* and work with a shorter survey. While a shorter survey tool will reveal less information, it may also result in less bias. Those who are willing sample volunteers for an abbreviated questionnaire may offer a much more accurate depiction of the true population.

The tolerance for the length of a survey varies with the method of the survey. For example, studies have found that people are more likely to accept a fairly lengthy mailed survey than participate in the same length survey administered over the telephone or in-person. The reason is convenience, which is our next subject.

"You want me to do this now?"

The telephone rings just as you are about to step into the shower. The doorbell rings just as you are about to leave for an errand. Someone intercepts you at the shopping mall while you are carrying two shopping bags. Upon returning home, your mailbox is stuffed with two dozen pieces of mail. You turn on your PC and discover within your E-mails that you are on *another* database mailing list. Weeks ago, you agreed to travel to a local auditorium to participate in a music preference study but it's snowing and you are tired.

Warning: Beware of surveys that are inconvenient in terms of timing and competing activities.

There can be dozens of reasons why the scheduling of a research study is inconvenient. As with the prior warnings, inconvenience can introduce sample bias. For many situations, soliciting more people ("oversampling") is the easiest solution, but suppose that no matter how many additional solicitations you make, a specific demographic group is still overlooked. For example, no matter how many phone calls are made during the day, most working women will not be available to answer. To rectify the problem, the

Troubleshooting Audience Research

researcher must call at night or over a weekend. Similarly, the day and time of in-person mall intercepts may have a profound influence on who is available to be in a sample. Properly trained telephone interviewers will alternate times and days in order to overcome no answers and busy signals.

A distinct advantage of mailed surveys is that the recipient can find the time when it is convenient to complete the survey. The obvious handicap for the researcher is the loss of immediacy and the considerable lag time from when the survey is mailed to when it is finally returned. Even with this positive convenience factor, mailed surveys, in general, have much lower cooperation rates compared to telephone methods. Telephone solicitations may have a reputation for annoying people, but over the long run, this technique still reveals better cooperation than the postal route.

"I started the thing but then quit...I just couldn't figure out what they wanted"

Earlier in our section on questionnaire design, we mentioned how poorly worded instructions and questions can lead to misleading results. While this confusion is bad enough, we now see another related problem where the participant simply *drops out* of the study.

Warning: Beware of surveys that have difficult instructions and overly complex questions.

Mailed surveys are particularly prone to this type of failure. Unlike telephone and in-person interviews, the participant cannot request help in resolving the confusion and since there is seldom any penalty for not completing the project, these frustrated people find something better to do with their time. Also, complicated open-ended questions, where participants are required to write long sentences, encourage people to walk away from a survey. A researcher should feel miserable with this kind of abandonment. A good metaphor for it is "the fish that got away."

Follow Through

We can see by the above reference to a fish story that once you have a potential participant "hooked," the battle still isn't over. It is vital to follow through with cooperation strategies during the administration of a questionnaire and the collection of the data. The cardinal sin is to recruit good sample participants and then have them go AWOL. Aside from the above-mentioned items, such as the length and complexity of the questionnaire, there are other cooperation factors that need attention.

First, the *physical surroundings* should be comfortable - comfortable temperature, comfortable lighting, comfortable seating, comfortable acoustics, etc. Second, the persons administering the questionnaire, whether in-person or over the telephone, should be *courteous and helpful*. Remember that the perceived payoff or reward for cooperating in this venture is usually psychological, meaning that the *personalities* of the people

administering the survey can be an important factor in maintaining cooperation. Survey participants should be respected and appreciated for their efforts. Finally, *keep your promises* about important aspects of the study, such as the promised subject matter, the promised length of the survey, and most importantly, the promised anonymity.

As with soliciting participants, the decision-maker may not be personally involved in administering a questionnaire, but this does not preclude an unscheduled “visit” to the research location to monitor the quality of work performed by the designated researcher. This can take the form of eavesdropping on telephone solicitations or conducting informal “exit polls” of persons leaving an auditorium or testing center. Imagine in our case study that the General Manager stops by the telephone market research headquarters and hears the following, “Lady, how long am I supposed to wait for you to make up your mind? Jeez!” Or suppose the GM queried a person leaving a mall survey site and heard, “The questions were stupid and the people in there are obnoxious!” Aside from being shocked personally at the way respondents are being treated, there is the additional consternation over the *quality of the research data collected*. Could the responses to some questions have been contaminated by the manner in which the participant was treated? This is the kind of sample bias that random sampling and statistical weighting cannot control.

Troubleshooting Audience Research

Chapter 5

Troubleshooting the Interpretation

At the beginning of this book, we emphasized the notion of a chain of events that leads ultimately to a meaningful study. No matter how conscientious a decision-maker and researcher are in analyzing the collected data, if the initial design and execution of the study are seriously flawed, there is no point in wasting time analyzing the numbers. The data may be so “contaminated” by poor problem definitions, poor questions, poor sampling techniques, poor cooperation rates and poor questionnaire administration that the best statistical analysis will not produce worthwhile knowledge. It’s too little too late.

Assuming that the study was planned and executed properly - and these are big assumptions - there remains one last challenge: interpreting the results.

THE DEVIL IS IN THE DETAILS, PART II: TROUBLESHOOTING THE FINAL REPORT

As with the presentation of a research proposal, the decision-maker should have *expectations* about the presentation of the final report. The number one consideration should be *clarity* - does the report make sense? The decision-maker should not be forced to translate esoteric research jargon and dense statistical analyses. A professional researcher should be trained in providing information for two types of readers:

- Management decision-makers
- Fellow researchers

The first major sections of the report should address the management decision-makers. Highly technical information, geared more for people with professional training in research methods and statistics, should be provided in footnotes or appendices.

Warning: Beware of incomplete or poorly organized final reports. Everything should have its proper place within the document. Missing elements and poor organization can lead to unnecessary confusion and frustration.

Here are some items that one should expect to see in a typical business research report.

Cover Letter or Memo from the Researcher

A cover letter or memo permits *personal communication* between the researcher and

Troubleshooting Audience Research

decision-maker that may not be appropriate within the formal written document. In addition to the usual courtesies, there may be proprietary comments that the decision-maker may not want to share with every person who will read the final report. For example, comments about research expenses and consulting fees should be confined to a cover letter and not included within the text of the report.

Report Cover Page, Table of Contents and Introduction

Remember that this report may be scrutinized by people who are not as familiar with the study as the management decision-maker who evaluated the original research proposal and supervised the study from beginning to end. Any media executive should be able to pick up your report and understand readily (a) what the study it is about, and (b) how the report is organized.

The cover page should include a simple yet well-crafted title that captures the core *purpose* of the study, not the final conclusions of the study. Conclusions may be a matter of interpretation and controversy. Annoying a reader on the very first page of a report is not the way to win the hearts and minds of a media management team! Regardless of the subject matter, the title should appear objective and impartial. Earlier, in the section dedicated to the research proposal, we focused on the need to eliminate hidden agendas and preconceived results. The final published report should also “rise above” the temptations of taking a position too early.

Bad Title: *Why Poor Weather Reporting is Damaging the Early Morning Newscast on WXXX*

Better Title: *An Analysis of Audience Perceptions of the WXXX Early Morning Newscast*

In addition to an appropriate title, the cover page should include the name of the designated researcher or research firm and for whom the study was conducted. See the following example:

An Analysis of Audience Perceptions of The WXXX Early Morning Newscast

Conducted by
Frank Marvel Associates
for
WXXX-TV
March 15, 2000

The table of contents is intended for the reader's convenience, not the researcher's utility. If the decision-maker who is responsible ultimately for the published report finds the table inadequate or confusing, he or she has every right to request a revision.

The intention of an introduction is to set the stage for the study. What was the problem that required more information? Who was commissioned to conduct the study? In general, what type of study was conducted? Often, an Introduction becomes part of our next topic, an Executive Summary.

Executive Summary

An Executive Summary is similar to what academic researchers call an abstract. As the term implies, it is a summary of the entire project intended to be read by an executive, not necessarily a fellow researcher. Unlike an abstract, an Executive Summary dispenses with research jargon and provides information in a more readable context. Because this may be the only section read by many company executives, it is imperative that this summary be well written (the original decision-maker may wish to see a draft before it is released as a formal document). In addition to a concise overview of how the study was planned and executed, the summary should include interpretations and conclusions. But these final two items within the summary can be troublesome. With the goal of *condensing information* for a busy, impatient reader, interpretations and conclusions can be oversimplified. Again, the supervising decision-maker may be aware of all the complexities of the study but others within the company may not have this insight. An Executive Summary should be written with this broader readership in mind. In fact, there may be situations where the *decision-maker*, rather than the researcher, writes this important section of the document.

Background

A Background section addresses what was *already known* about the situation before the study was conducted. This should include a more detailed description of the research problem (including data) and pertinent information from other research sources, such as similar studies conducted in other markets. Using our case study, we would present as background all the known ratings data and audience flow studies acquired from Nielsen.

Description of Method

The Description of Method section should provide the reader with a clear understanding of how, when and where the study was executed. In addition to noting the target population under study, the researcher should present the *sampling techniques* used to replicate this population. Furthermore, a copy of the questionnaire should be included in this section or filed as an appendix. If this questionnaire is an "original" and not borrowed from another study, the researcher should provide information derived from focus groups and pilot-testing that contributed to its construction. Highly trained professional and academic researchers would use the terms *reliability* and *validity* to justify the use of a particular questionnaire. Also, a competent and ethical researcher should present *cooperation rates* in this methods section.

Troubleshooting Audience Research

Results

The Results section is the heart of the final research report. Depending on the sophistication of the study and its intended readers, certain statistical information may be diverted to an appendix, but all data should be available *somewhere* within the report, especially if the study has the potential of stimulating controversy. Denying access to the “nuts and bolts” data can only aggravate a volatile situation.

The researcher should provide ample charts, tables and graphs to bring the raw data to life and to make the statistics more comprehensible. However, as the next section of this book explores, beware of “smoke and mirrors” that distort or disguise information. Results should be presented in the most *revealing yet impartial* manner possible. A researcher’s personal point of view should not be known in the Results section. In a perfect world, one could assert that a *substitute* researcher should be able to step in and offer the same results without the reader knowing of a personnel change. Results are results.

Conclusions and Recommendations

This final portion of the research report allows the researcher some personal or subjective input. Depending on the study, some conclusions may be so obvious that any knowledgeable reader would agree. But more often than not, conclusions are a matter of gray scale rather than pure black and white. It is vital for the integrity of the study that a clear distinction exists between results and conclusions. Better yet, another section entitled *Conclusions and Recommendations* can further differentiate fact from opinion. Going back to our case study, we could have a situation where the statistical analysis of responses generates an obvious, uncontested conclusion that one of the program segments is weak and in need of repair, but the nature of this “repair” may not be so obvious. There could be several *recommended* solutions from which to choose - some provided by the designated researcher and some provided by members of the management team. In a more contentious situation, the statistical results may support a conclusion *but only to a limited degree*. Academic researchers are trained to never say results “prove” anything. At best, the results “support” a hypothesis or “suggest” that something has a “high probability” of being true. In social science research, of which audience research is a part, few things are absolute.

The Recommendations and Conclusion section can allow a researcher to bring to bear the full extent of his or her expertise. Often, a highly trained and experienced eye is required to uncover complex information from a statistical analysis. In particular, a competent researcher will administer “significance tests” on the raw data to determine whether the results are genuine or merely an illusion manifested by random sampling error.

Oral Report

The designated researcher should be expected to present an oral report to an appropriate audience of decision-makers and possibly other researchers.

Ask the Researcher to prepare an oral presentation of the study.

This is an opportunity for the researcher to field questions and clarify issues. In order to enhance the quality of discussion, the audience should receive the written report *prior* to this presentation. This prior exposure should stimulate better questions. In the beginning of this handbook, we recommended that the researcher be protected from internal office politics and hidden agendas. During this oral presentation, top management should still offer suitable protection but the researcher should not be shielded from *legitimate criticism*. Just as most media managers are held accountable for their actions in various ways, so the researcher should be “held to the fire.” Comments and criticism should address the *substance* of the study and avoid pointless personal attacks.

REALITY OR SMOKE AND MIRRORS? - THE ART AND SCIENCE OF STATISTICAL DECEPTION

Obviously, Mark Twain had little respect for those who toil with numbers when he accused statistics of being worse than “damned lies,” and some of his disdain probably is justified. For decades statistics have been manipulated for both good and evil causes. Perhaps this use of statistics to justify an already established belief is the root of the problem. But statistics themselves don’t deserve such indictments. When used properly, they can provide a better understanding of reality. When abused, they become associated with the “smoke and mirrors” used by people who want to disguise or distort reality.

Media decision-makers encounter statistics in two ways. First, these managers are bombarded with advertisements, brochures and presentations that utilize statistics to help sell something:

- “Nine out of ten News Directors prefer this digital video camera.”
- “This is America’s number one off-network first year daily- strip family syndicated drama airing in early fringe!”
- “Our cable network has grown by 500% in one year!”

The second situation where a manager encounters research statistics is when he or she deals with syndicated or proprietary research. This book has focused on the second encounter but the media decision-maker should also be a savvy consumer of research-based sales pitches. The following discussion will help in both areas.

Averages and Variances

We use averaging to simplify large data sets by generating a single statistic that is the typical or most representative score. From determining automobile gas mileage to predicting annual rainfall, we are surrounded with averages in our everyday lives. From a more professional perspective, media executives calculate and manipulate averages to make dozens of common management decisions. Nielsen and Arbitron reports are filled

Troubleshooting Audience Research

with references to averaging - Average Quarter Hour (AQH) is the bedrock unit of analysis for most ratings measures. HUTs, ratings, shares, and cumes all include an element of averaging. Researchers will often use the term *mean* to denote *average*. In mathematical terms, a mean is calculated by adding all the values in a set of numbers and dividing by the number of values. The *median* is another statistic often used to depict a typical score but instead of taking the sum of all the values, the median requires that all the values be arranged in ascending order. The actual median is the exact middle score of this progression. In many cases, the median is a good statistic because it is not as sensitive to one or two extreme scores in a set of numbers as the mean can be. But most statistical analyses require a mean score to execute the formulas properly.

Warning: Averages and means can be misleading. Beware of research reports that do not present statistics describing the variance surrounding average scores.

While averaging is a convenient way to summarize raw data, there is a downside. In addition to calculating the typical score on a survey, we also need to look at the spread of numbers about that score. This spread is often called *variance*. When the degree of variance is substantial, an average or mean score can be a misleading statistic. Suppose the sequence of numbers presented below represent the scores on a personality assessment of morning news anchors where 1 represents “thoroughly obnoxious” and 5 represents “thoroughly delightful.” Let’s say that ten people participated in the study.

Personality A data set 1, 1, 2, 2, 3, 3, 4, 4, 5, 5

Personality B data set 3, 3, 3, 3, 3, 3, 3, 3, 3, 3

Personality C data set 1, 1, 1, 1, 1, 5, 5, 5, 5, 5

Personalities A, B and C exhibit the same average score of 3.0, yet the variances of the three data sets are radically different. Version A shows more of an even distribution across all possible scores, implying that he has a broad range of appeal. Personality B has a “middle of the road” personality where nobody has any extreme feelings in either direction. Personality C, however, exhibits a classic “polarizing” personality where people either love him or hate him.

In the above case, the degree of variance implies different management decisions. Is it better to have an on air personality that is adored by some viewers but hated by others, or is it more prudent to keep an anchor that is fairly innocuous and doesn’t excite or offend anybody? Imagine being unaware of these variances and making an important decision, such as a contract renewal, based solely on the average or mean score!

Standard Deviation

Knowing how far from the center data tends to range can be valuable information. The most common statistical measure of this spread is *standard deviation*. It is not necessary to

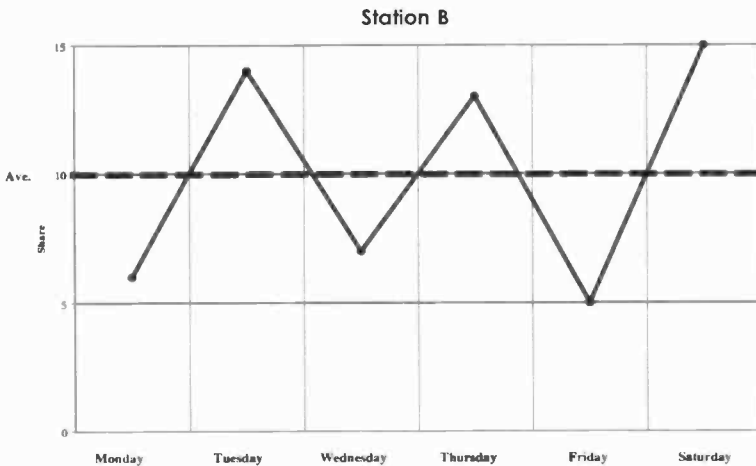
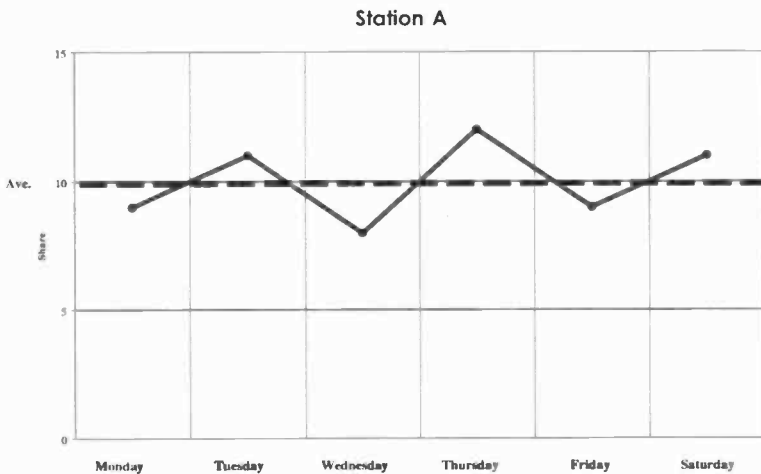
memorize the algebraic formula for standard deviation, but it is worthwhile to understand conceptually what the statistic represents. Think of standard deviation as a measure of the average distance of the data from the mean. The bigger the standard deviation, the less reliable the mean becomes in predicting certain outcomes.

Imagine the data set below represents the number of hours per week each of ten people spends using the Internet.

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

The average or mean score is 5.5, but how many people from this group actually spend 5.5 hours using the Internet? From a mathematical perspective 5.5 is the most “typical” score but with so much *variance*, does this statistic offer much insight?

Similarly, when comparing the long-term average performance of two or more competitors, variance (standard deviation) can be an important statistic. The two graphs below depict the seven-day performance of two stations. Notice that while both stations



Troubleshooting Audience Research

deliver the identical *average share* of audience (the dotted line), station A exhibits far less variance (smaller standard deviation) than station B exhibits. The “peaks” and “valleys” are not as dramatic.

Rankings

Everybody wants to be ranked “Number One,” so this type of statistical analysis is common within the broadcasting industry. Used as a sales tool, rankings are harmless, but as a media management tool rankings can disguise important information.

Warning: Beware of rankings that make no reference to the original data scores.

The problem lies in the interval between the rankings. When looking at a list of rankings, such as the example below, the mind presumes automatically that the mathematical distance (or interval) between the rankings is the same. But as anybody who follows horseracing knows all too well, the distances between win, place, and show can be by a nose or a quarter mile. In our example the ratings interval between #1 and #2 is only half a rating point, but the interval between #2 and #3 is six and a half points. Now let's factor in some sampling error of 1 point and you have a statistical *tie* for first place.

Radio Rankings/Ratings Example

Program	Ranking	Rating
A	1	11
B	2	10.5
C	3	6
D	4	5.5

The management team at station A should not take their “Number One” position too seriously. First, its nearest competitor is “breathing down its neck.” Furthermore, the majority of listeners (66%) in this 4-station market do not listen to the “Number One” station. In a formal research report, rankings are valuable for organizing data but should always be accompanied by the actual data.

Another problem in interpreting rankings is derived from questionnaire design. As noted earlier, people may feel the same about two items but are forced to rank them differently by the survey instrument. The reported rankings thus become somewhat artificial.

Percent Change

Percentages are used when a researcher wants to make comparisons. Instead of stating that station A delivered a 15 rating and station B delivered a 10 rating, we translate the

two items of data into a single percentage - Station A delivers 50 percent more audience than station B.

Warning: Be wary of comparisons of change that use percentages. *Switching bases* can disguise or exaggerate real change.

Converting data into percentages or proportions seems innocent enough until we want to compare change. The following table illustrates how percentages can deceive.

One-Year Comparison of Ratings

Station	Last Year	This Year	Change	% change
A	20	15	-5.0	-25%
B	5	10	+5.0	+100%

Despite a 100 percent increase in ratings, which of the two stations delivers the highest ratings? This is a common tactic used in advertising, but it is also an interpretation device used in formal research reports. The problem is *switching bases*, meaning that the comparison is not based on a common mathematical base but rather on two or more bases. Each of the above stations changed 5 rating points, but the percentages were calculated against two different bases. This is an ideal example of an “apples to oranges” abuse of statistics.

Looking at our case study, we could have a situation where a competitor is generating sales presentations bragging that its morning newscast has increased audience size by a huge percentage. The trick is that the station began with lousy ratings and therefore *any* increase looks impressive when converted to a percentage! This type of deception is tolerated in advertising and promotion campaigns, but real management decisions should be based on less inflammatory statistics.

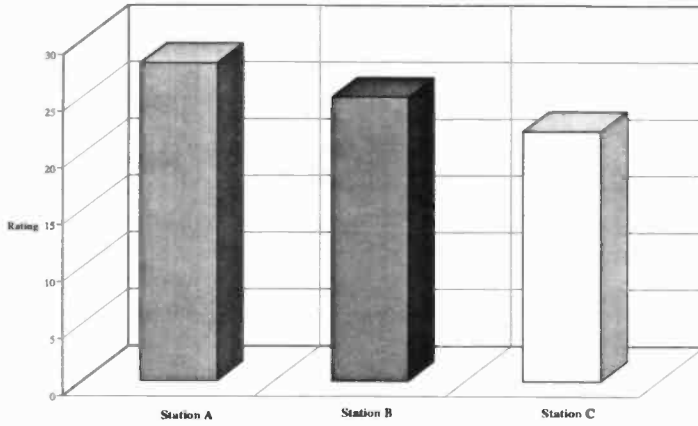
Optical Illusions

Charts and graphs and other types of illustrations should be used to make statistical data more understandable. However, they can also be exploited to exaggerate results. The most common deception is to exaggerate variance by not beginning a graph at a true zero point. On the following page is an example of how a bar graph can accentuate the variance or disparity among variables. Graph A begins at zero while Graph B begins at 20. While the mathematical separation in the stations’ ratings is identical for both graphs, the second graph gives an optical illusion of greater disparity among the three stations.

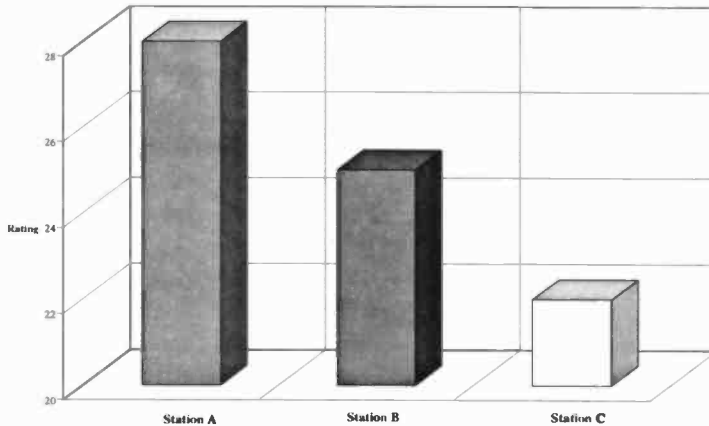
This graphic deception is somewhat similar to our discussion of switching bases when dealing with averages. The data are not necessarily wrong, but are susceptible to misinterpretation.

Troubleshooting Audience Research

Graph A



Graph B

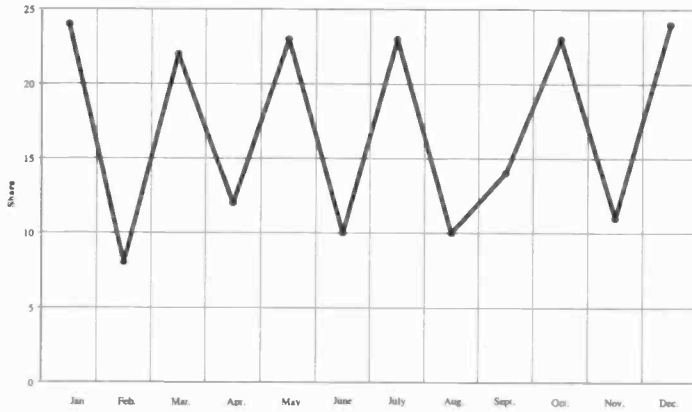


Another example of an optical illusion is in the *choice of scales* used to demonstrate variability over time. On the following page are two graphs depicting the same audience trend over time and yet Graph A gives the impression of greater fluctuations than Graph B does. This perceived difference is a result of the scale selected for the vertical axis on the left side. Notice how the range of numbers varies between the two presentations. Graph B compares a program's share performance against a scale of a full 100 percentage points. On the other hand, Graph A uses a range of only 25 percentage points to depict *the same fluctuations* in share. Depending on the agenda of the researcher, the relative stability of the station over time can be distorted graphically in either direction.

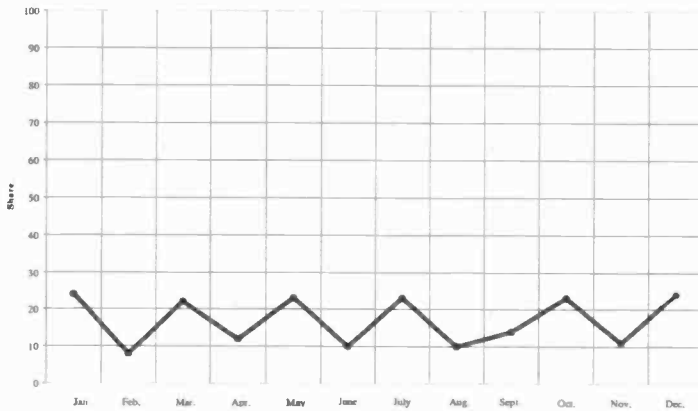
Correlation vs. Causation

The interpretation of statistical information can be a source of debate among managers. For example, many surveys generate statistical correlations where two or more factors (variables) are analyzed to see how closely associated they are.

Graph A



Graph B



Warning: Evidence of a close association (correlation) between two factors does not necessarily mean that one causes the other to change.

Problems arise when a close association is interpreted as causation - one factor is presumed to cause a change in the other. Let's take an example from nature. There is a strong correlation between the time when the sun rises and the first cries of a rooster in the morning. However, nobody should conclude that the early morning crowing causes the sun to rise. While this example might seem silly, there are similar situations in quantitative media research where the silliness is not as easily recognizable. A high correlation between (a) a newscast's ratings and (b) the day of the week does not necessarily mean that audiences prefer the content of one night's newscast over another. The genuine cause of one night's exceptional ratings performance may be an unusual lead-in program or a "run over" late-start for a competing newscast or the addition of a highly promoted weekly feature.

Troubleshooting Audience Research

Because two sets of numbers are observed to correlate, or increase or decrease simultaneously, it does not follow that one set caused the other to happen. In most cases, it's a coincidence. Professional and academic researchers use the term *confounding* to describe situations where there can be two or more plausible explanations for a given statistical outcome.

In the above example, there are several confounding variables that may or may not have caused an increase in household ratings. Only a controlled experiment can offer the proper insight concerning true cause and effect relationships.

Reporting Sampling Error

Another common problem in interpreting statistical results is the failure to recognize the statistical margin of error found in any sample-based research.

Warning: Beware of "absolute" answers derived from sample-based research. As long as samples are used to represent a larger population (even perfect random samples), the results will still be estimates that contain sampling error.

"Sampling error" or "margins of error" are expressions of *reliability*. Sample reliability is analyzed using two statistics. Standard error is the plus or minus (+ or -) degree of error that can be expected from an estimate. For example, suppose a radio station wanted to know on average how many hours a week a teenager listens to the radio. The station manager commissions a study using a large random sample. The sample estimate was an average of 10 hours per week with a standard error of plus or minus 2 hours. This means that the manager can be confident to some extent that the true measure of the parent population lies somewhere between 8 and 12 hours. This range (or interval) of numbers is called a *confidence interval*.

Associated with standard error and confidence intervals is the notion of confidence level, which is a percentage statement of *how confident* the researcher can be that the sampling error calculations are true. Another way of looking at this term is to ask the question, "If an infinite number of samples were drawn, how confident would we be that the sample estimates would all fall within this confidence interval?" The laws of probability coupled with some algebra will reveal that one can only be 68% confident that the true population measure is found within one standard error. Doubling the standard error statistic increases the confidence level to 95%. Returning to our example, we can say that we are 95 percent confident the true population measure of the average number of hours spent listening to the radio lies somewhere between 6 and 14 hours.

A station manager, wanting a really precise answer, may be disappointed in these results. The calculated sampling error makes the original estimate of 10 hours seem a little suspect, but this is the reality of sample-based research.

As discussed earlier, one standard error unit, by definition, delivers only a 68 percent level of confidence. The reason why is beyond the scope of this handbook, but an introductory statistics textbook can enlighten the more interested reader. For our purposes here, we are asking the reader to take a blind “leap of faith” and just accept the fact. A 68 percent confidence level means that if we took 100 random samples of this population, we could expect that 68 estimates would fall within the specified confidence interval. The remaining 32 sample estimates would fall outside this interval, meaning that it is still possible that the true population value is also outside this interval. By widening our interval by two standard error units, we increase our confidence to 95 percent but there is an obvious trade-off: *less precision*. Ninety five percent confidence (2 standard error units) often reveals intervals that are so broad that it is impossible to make a prudent management decision.

Ask the Researcher: What is the confidence level of the reported sampling error (standard error or confidence interval)?

Because tiny samples generate high sampling error, managers should not agonize over tiny differences in program or station performance. This is particularly true with single digit Nielsen and Arbitron ratings. In the front of every “sweep” book, the ratings services provide some basic guidelines about applying standard error to their audience estimates. More detail will be provided in a later section. It is important to remember that both Nielsen and Arbitron admit to considerable sampling error in their own published reports. For research other than ratings, the same warning applies.

Warning: Small estimates taken from small samples tend to generate big sampling error.

A formal report depicting all the statistical results of a study may contain dozens of impressive tables and dazzling graphics but the decision-maker must remain skeptical of the data. A good researcher would offer sampling error estimates along with the “hard” data presented. Regrettably, sampling error is often not discussed because the researcher is worried that the study will appear weak and inconclusive. This is both foolish and dangerous. The decision-maker who commissions a study should understand the inherent statistical limitations of sample-based research and insist on their disclosure in the final report. The example below illustrates how the refusal to admit sampling error can jeopardize a management decision by giving a false impression of *exactness*.

A radio station commissions an audience study that examines the popularity of certain talk show topics. The management of the station desires its talk show hosts to concentrate on topics that will have the broadest appeal. Let’s presume that the study was well designed and well executed and the statistical results revealed the following.

Troubleshooting Audience Research

Four Most Mentioned Topics

<u>Topic</u>	<u>% mentioned as #1 choice</u>
Sex	25
Money	20
Health	15
Politics	10

At first glance, there seems to be some significant differences among all the topics, but now let us introduce a standard error of plus or minus 5 percentage points. Adding and subtracting these points from each percentage reveals considerable overlap. The sampling error *confidence intervals* are “saying” that we can’t be sure of the exact ranking. We can be 68 percent confident that the true population value falls within a range of five points on either side of each percentage estimate. Doubling the standard error to plus or minus 10 increases our confidence level to 95 percent, but this range (or interval) of numbers is now largely meaningless for making an informed management decision.

Chapter 6

Troubleshooting Nielsen and Arbitron

For good or bad, ratings are a fact of life for broadcasters. In the business world of buying and selling audiences, these numbers are the tools of transaction. The relationship between broadcasters and the Arbitron and Nielsen ratings services have not always been marriages made in heaven. For decades, networks and stations have accused these audience research companies of shoddy work, and yet when a formal complaint is filed, Nielsen and Arbitron are usually vindicated. The reason, in most cases, is that the plaintiff fails to recognize the inherent limitations of sample-based surveys. These companies publish detailed information concerning these limitations in their standard reports and in annual methodology supplements made available to all interested clients and researchers. While some sections may be difficult to comprehend, media decision-makers should attempt to read these publications, especially if their station or network wishes to object formally to the findings of a survey. Furthermore, these published limitations should be “must reading” for any in-house research director or hired consultant.

Obviously, most complaints stem from disappointing ratings (what broadcaster would complain about high ratings!). The storyline runs like this - When reported audience ratings are high, it's the result of brilliant programming and marketing. On the other hand, when the reported audience levels are low, it's the result of bad research procedures.

Decision-makers must understand that most of the blame for poor ratings is not due to incompetence but to chance. This means that high ratings often can be just as imprecise as low ratings. Earlier we learned that sample-based research deals with the laws of probability in that samples generate *estimates*, and small estimates are more prone to error than large estimates. Additionally, sampling error calculations are based on the impossible assumption of a mathematically perfect random sample.

READ THE FINE PRINT

At the beginning and end of all Nielsen and Arbitron sweep reports are several pages written in incredibly small type. Let's take a magnifying glass to some of these statements.

Warning: Beware of battles you can't win. Ratings are merely estimates. The ratings services, in their published reports, admit to imperfect samples and considerable sampling error.

Troubleshooting Audience Research

The following quotations are taken directly from Nielsen and Arbitron published reports.

From Nielsen

"The use of mathematical terms to express the audience estimates therein should not be regarded as a representation by Nielsen TV that they are exact to the precise mathematical values stated."

"Since the audience estimates in this (report) are obtained from a sample, they may differ from estimates based on a complete census of TV households in the same sampling frame and using the same methodology....Standard error is a measure of sampling variability for a probability sample. The standard errors apply only to a perfect probability sample. The achieved sample is not a perfect probability sample primarily because some households do not cooperate."

"The responsibility for use of information based on samples, including judging the magnitude of errors and recognizing situations where differences are of no statistical significance, lies with the user of such information."

From Arbitron

"Arbitron estimates are subject to statistical variances associated with all surveys which use a sample of the universe.....Users of this report should keep in mind that, due to (these) factors, the reliability of Arbitron estimates, data and reports and their statistical evaluators cannot be determined to any precise mathematical value of definition."

"The reliability of a survey estimate is generally defined in terms of the estimated margin of sampling error around the estimate, or confidence interval. The confidence interval describes the extent to which a survey estimate, based on a random sample of a population, may vary from the result that would be obtained through a complete census of that population."

The majority of complaints against these firms concern an unexpected drop in the ratings performance of a program or daypart. Nielsen and Arbitron defend themselves usually by analyzing the sampling error (standard error and confidence intervals) based on (a) sample size, (b) demographic composition, and (c) the variability of responses. The result is often the identification of what researchers call statistical "bounce" which is a fluctuation in ratings caused by random error rather than true audience behavior. Remember, standard error is expressed with a (+) sign and a (-) sign, implying that this "bounce" can *inflate* as well as *deflate* ratings. That is, a perceived increase in estimated audience size can be as unreliable as a perceived decrease in audience size. Standard error is a two-way street! Depending on the degree of sampling error, what may appear to be a close race among several competing stations can, in reality, be statistically meaningless.

CALCULATING SAMPLING ERROR

Each ratings company approaches the calculation of standard error in a different manner.

Nielsen Diaries

In the front of a typical Nielsen Station Index sweep report are several pages dedicated to estimating standard error. These paragraphs may be difficult reading but the content is worth the effort (if an NSI book is available, the reader is encouraged to grab it and follow along here). Rather than provide standard error data for each rating, which would be impractical, Nielsen provides a couple of ways to gauge approximate error.

A section entitled "STANDARD ERRORS: FOUR-WEEK AUDIENCE ESTIMATES" provides 5-day and individual day standard error statistics for Household and Persons ratings of 1, 5, 10, 15, 20, 35 and 50. Regrettably, the sampling error for other possible ratings (such as a 7 or a 12) cannot be estimated without going through some algebra. The formulas and step by step procedures for doing so are presented in the annual Nielsen Methodology Supplement.

The standard error estimates are expressed in terms of rating points and in total audience. For example, a 5 rating over a 5-day average (e.g. a Monday through Friday strip program) reveals a standard error of .2 or a confidence interval of 4.8 to 5.2. Similarly, a one rating in a particular market may translate into 10,000 households. Let's pretend that the table indicates a 5-day average standard error of 1,500 households. This statistic means that this estimate would have a *confidence interval* of 8,500 to 11,500 households.

The tables are cumbersome to use, the instructions are difficult to comprehend, and the type is far too tiny to read comfortably. However, Nielsen does make an attempt to provide some information on sampling error. Another section entitled "RELATIVE STANDARD ERROR (1 Standard Error)" explains a different way to look at sampling error by expressing the *standard error* as a percentage of the ratings estimate. For example, a program that delivers a 10 rating with a standard error of 1 would generate a relative standard error of 10 percent (one divided by 10). The section then explains how at the beginning of key sections inside the report there are estimates of Relative Standard Error Thresholds ("R.S.E. Thresholds") expressed as two percentages: 25% and 50%. Across from each percentage are several columns of ratings referring to households or age/gender demographic categories. The numbers represent the ratings threshold whereby any rating in the actual report that is equal to or less than that threshold number will exhibit at least a 25% or a 50% relative standard error. For instance, let's assume that in one demographic column (men 18-34), the 50% R.S.E threshold is a 2 rating. Now let's propose that a particular program delivers a 1 rating within this demographic. Because this rating is less than a 2, Nielsen is saying that we can count on sampling error of at least 50% of the estimate, meaning that the standard error is roughly plus or minus .5 rating (50 % of 1) or a confidence interval .5 to 1.5. The same calculations could be made at the 25% threshold level. Nielsen states that "Audiences below these (threshold) levels are of marginal statistical significance." For most TV markets, audience estimates of one or two ratings points are meaningless and should not be used to make important management decisions.

Troubleshooting Audience Research

Nielsen Metered Markets

For TV stations in Nielsen's metered markets, sampling error information is also available. The company merges the metered household data with demographic information derived from the periodic diary-driven sweeps. The standard error approximations are based on this hybrid sample base.

Nielsen has begun the process of replacing the current passive household meters used in local markets with its more sophisticated People Meters (similar to the type of meter used in its national ratings sample). These new devices will report demographic data and therefore may eventually eliminate the need for diary-driven sweep periods altogether. Many buyers and sellers of TV audiences welcome the changeover.

Because electronic meters are expensive to install and maintain, the sample size for metered households is significantly smaller than the diary sample used for the periodic sweep periods. As we already know, small samples generate more sampling error than do big samples. Also, participating metered families will remain active in the sample for years whereas a diary-based household participates for only one week. With some homes remaining in the panel for as long as five years, a serious bias within the metered sample is more likely to linger and contaminate results for a much longer time than it would with the one-week turnover diary method.

The current systems record tuning behavior every 30 seconds, and by special request Nielsen will provide minute by minute tracking of a program's performance. However, broadcasters should be extremely cautious in using these data. The standard error "bounce" effect can be huge compared to weekly averages of total program performance. Snap decisions should not be made based on what might be highly unreliable ratings data.

Returning to our case study, the management team will drive themselves crazy if they track the newscast's performance on such a tiny scale. *Long-term trending* provides the most reliable sample estimates. The General Manager should discourage impulsive evaluations of any changes to the program. It not only takes time for audiences to become aware of changes, but also it takes time for ratings that are influenced by sampling error to "level out."

Arbitron Diaries

Unlike Nielsen's broad approximations, Arbitron gives the reader a series of tables and mathematical formulas that enable the radio broadcaster to calculate precise standard error statistics. The specific section in the report is called "ARBITRON RADIO RELIABILITY TABLES" and includes well-written instructions and examples (again, if an Arbitron book is handy, grab it and follow along). The procedure is not terribly difficult to understand. All you need is a hand-held calculator and some patience. Here is an example. Suppose a particular radio daypart delivers a 5 rating. You go to "Table A" and directly across from a 5 rating is a special number: 21.79 (never mind what it means). Next, you go to table B to match (a) the desired demographic group with (b) the desired daypart. In our case we want men 18 to 34 and morning drive time. Again, a special

number will be revealed where these two factors converge on the grid: 41.34. Finally, divide the special number from table A by the special number from table B and bingo, you have the standard error! In our example, 21.79 divided by 41.34 equals .53 (plus or minus). Our rounded off confidence interval would be 4.5 to 5.5. As with Nielsen, the Arbitron Company can provide even more detailed statistical information on sampling error.

Remember that one standard error yields only a 68 percent level of confidence. The data is saying that if you took many more random samples of the same population, roughly two out of three times the sample estimates would fall somewhere within your calculated interval. Another third of the estimates would fall outside this range of estimates. The odds are that the true population value is contained within your "two-thirds" interval, but there is no guarantee. To better your odds, you can double your standard error and recalculate a much wider confidence interval. Your confidence level will rise to 95 percent but the precision of your estimate has been reduced considerably. Here is a quick example.

Arbitron estimates that your radio station in morning drive delivers 10,000 adults age 18 to 49. The calculated standard error (plus or minus) is 1,000 persons, yielding a 68% confidence interval of 9,000 to 11,000 persons, but is this level of confidence acceptable? Dare you make an important programming decision based on these odds? You elect to go to a higher level of confidence and double the standard error, yielding a 95% confidence interval of 8,000 to 12,000 persons. Feeling any better? Not exactly. You are more assured that the true population value lies within this range of numbers, but now the range is so wide (a difference of 4,000 persons), you still can't make a well-informed decision!

SAMPLE COMPOSITION

While recognizing survey limitations due to the mathematics of sampling error, some broadcasters still accuse the ratings services of poor methodology regarding the *representativeness* of the sample. In these cases, the critics assert that the number of in-tab diaries either under-represent or over-represent a particular demographic group or geographic area.

Warning: Beware of battles you can't win, Part II. Remember that the ratings services apply statistical weighting to compensate for demographic shortfalls in sample composition.

Here again, the broadcaster needs to become familiar with the exact details before firing off a complaint about sample composition. When generating a sample, both Nielsen and Arbitron use data acquired from the U.S. Census Bureau and other respected organizations that estimate annual demographic changes. Using a multi-stage cluster sampling technique, the companies generally solicit participants by telephone and then

Troubleshooting Audience Research

follow up with a letter. The number of people represented by one in-tab diary is called Persons Per Diary Value or PPDV (Households would be HPPV). The ratings services will apply statistical weighting or "sample balancing" to compensate for situations where the in-tab sample is not representative of the parent population. Some of the factors that are weighted by Nielsen include the following:

- Geographical location
- Age of head of household
- Household size
- Presence of children in the home
- Hispanic or African American identity
- Cable status

Arbitron radio samples go through similar adjustments. In some regions of the country, the mathematical formula used for weighting certain disproportionately represented ethnic groups has been the source of much controversy between local stations and the ratings services.

The statistics used in these procedures are quite sophisticated, and therefore any legitimate criticism from a client subscriber will require the same mathematical depth. If a decision-maker suspects that a sample is not representative of the market population, he or she should first refer to the detailed statistical supplements published by the research firm and then retain the help of a professional researcher to analyze the data. Once the formulas for demographic weighting are disclosed and the standard error of audience estimates is analyzed, most complaints are dismissed. Remember that Nielsen and Arbitron admit in writing that the data they provide comes from *imperfect weighted samples* that exhibit *substantial sampling error*. It is the clients who want these numbers to be perfect representations of audience behavior. Unfortunately, survey research is not an exact science.

The obvious solutions to these imperfections are (a) bigger samples and (b) better cooperation rates and (c) more frequent respondent turnover (for metered homes). In recent years, both Nielsen and Arbitron have increased sample sizes, but remember that the payoff in increased reliability reaches a point of diminishing returns (see the earlier section on sampling techniques). Responding to the many complaints expressed by subscribers, both firms have launched numerous initiatives intended to increase cooperation rates.

WHY RATINGS ARE IMPERFECT

The following is a summary of reasons why radio and television ratings are imperfect measures of audience behavior. You will see that for many of these problems, the researcher has little control.

1. **Samples produce sampling error.** Based on the laws of probability, even a mathematically perfect random sample will still produce estimates that have some error.
2. **Small ratings are more unreliable than big ratings.** Not all ratings are equally reliable. The mathematics of probability reveal that the larger the estimate, the more precise it is in replicating the true population value. Therefore, programs generating small ratings are more vulnerable to random fluctuations or “bounce.” As television audiences become ever more fragmented, this issue will persist for years to come.
3. **Samples are inherently biased because not everyone has an equal chance of being selected.** Arbitron’s samples and most Nielsen panels are drawn from households with a telephone. Nontelephone households are thereby excluded. In addition, business establishments and group quarters are also excluded. This means that any viewing or listening done at a bar, motel room, college dormitory or military barracks will not be measured (in some vacation markets, such as Orlando, Florida, there have been some attempts by Nielsen to measure motel/hotel viewing).
4. **Of those approached to be in a sample, not everyone agrees to participate.** With cooperation rates dropping every year, 30 percent diary response rates are not atypical today, and some Nielsen markets even track below the 20 percent mark. Many researchers believe this is the most crucial issue in audience research today. Noncooperating persons may have viewing or listening habits that differ from those who do participate in a survey. Statistical weighting is used to offset demographic disparities in a sample, but now the researcher may be placing extraordinary weight on only a handful of responses.
5. **People have faulty memories.** As discussed earlier, even with the best of intentions, people often fail to recall accurately their media habits. Nielsen and Arbitron encourage diary keepers to enter data as soon as possible, but studies indicate that most participants wait until the end of the week before they even open the diary. In these situations, “top of mind” recall becomes a driving factor, where more familiar programs are recalled more readily than less popular program items.
6. **Some people lie.** Sometimes diary keepers deliberately don’t tell the truth. They may be embarrassed to disclose some of their viewing preferences or they may report viewing that did not really occur. Studies have found that participants will often log in a favorite program title even though they were unavailable to watch during the designated survey week.
7. **Even meters have problems.** While many of the problems involving inaccurate diary keeping are eliminated when electronic meters are introduced, these devices are not perfect. Just because the TV set is “on” does not mean necessarily that it is being watched by anybody. Both Nielsen and Arbitron are experimenting with some high tech “recognition” devices that will identify who is in a room and looking at the screen. Sample sizes and turnover of participating households are additional reliability concerns.
8. **Programming and promotion hype can distort ratings.** Media buyers, advertising agencies and audience researchers suspect that these activities distort what should be *typical* listening or viewing behavior.

IF YOU STILL WANT TO FILE A FORMAL COMPLAINT

Yes, it's free country and you have the right to complain about anything you want. If, after considering all the above mentioned warnings, a broadcaster still wants to issue a formal complaint with one of the rating services, then the author wishes you Godspeed.

Ask the Researcher: Do we really have a case or are we just wasting valuable time and resources?

Indeed, there are documented cases where Arbitron and Nielsen have admitted to sins of methodology - beyond mere sampling errors - that led to "recalls" of some market reports. Nobody is perfect all the time and there can be lapses in judgment even among the most preeminent audience researchers in the world. Probably the best troubleshooting device for these services would be *competition*. Right now these two companies hold monopolies in their respective media fields and seldom do monopolies nurture serious self-criticism and change. A company called Statistical Research, Inc. created a TV audience measurement system called S.M.A.R.T (Systems for Measuring and Reporting Television) in the mid 1990s in an attempt to provide a national ratings service to rival Nielsen's. But after several years of pilot testing and soliciting financial support, S.M.A.R.T folded before its scheduled debut. Arbitron has encountered several competitors over the years, including AccuRatings, but still maintains enormous clout among radio stations and advertising agencies. It should be noted that both Nielsen and Arbitron also have entered the *Internet measurement* arena, along with several other upstart companies, thereby attempting to secure their positions as media "scorekeepers" far into the coming digital age.

Concluding Remarks: Don't Give Up the Ship

After encountering so many *warnings* in this text, a broadcaster may become suspicious of any type of quantitative audience research, but this would be a mistake. Properly *designed, executed and interpreted* research can yield valuable knowledge. Don't give up the ship because of a few storm warnings! Relying on intuition or gut instinct to make important management decisions is an even more risky proposition, especially when your competition is seeking as much knowledge as possible about your audiences and your advertisers. For every example of poor research, one can also find a gem - a brilliant study that reveals new knowledge about the thoughts, feelings or actions of audiences. What distinguishes good research from bad research is often the quality of troubleshooting invested in the project. The notion of *anticipating what might go wrong* lies at the heart of good troubleshooting, and so the author hopes that this handbook will help the reader steer the right course through some of the rougher waters of audience research.

Troubleshooting Audience Research

Appendix A

SOLICITATION GUIDELINES AND SAMPLE LETTER FOR MAILED SURVEY

- Use an attractive and useful letterhead. Make sure that company names, street addresses, phone numbers, websites and E-mail boxes can be found easily. You want to appear as *accessible* as possible.
- Ask for a favor. From the very beginning, admit that you need this person and the tangible rewards you can offer are minimal or nonexistent.
- Explain the purpose of the study and why it is important. In some cases, the study may be important personally to the recipient
- Disclose how you found this person's name and address. People are sensitive about their privacy.
- Estimate the duration of the questionnaire in terms of *minutes*, not the number of questions.
- Present explicit instructions as to how to complete the questionnaire and how to mail it back to the researcher. Also, emphasize how *easy* the procedure is to execute.
- Offer anonymity and/or confidentiality. Because you have a person's name and address (and telephone number), you already have invaded their privacy to some degree. They want to be assured that this is as far as you will go.
- Present a deadline with a sense of urgency. Even if a person agrees in principle to cooperate, he or she may procrastinate for days.
If possible, offer to send a report on the results of the study. This could be the one factor that persuades a person to come on board. Of course, you do not want to reveal important information to your competitors.
- Be appreciative. Your costs here are nothing but a few kind words.

Troubleshooting Audience Research

Acme Audience Research Inc.
1546 West Main Street
Middletown, Connecticut 01234
(203) 555-1212
Website: www.Acmeresearch.com
e-mail: customerservice@Acme.com

Dear _____:

We need your help. Our company is conducting a survey concerning how people think and feel about local TV newscasts in Middletown. In particular, we are interested in your honest views on early morning programs that are telecast between 6:00 and 7:00 am. We value your opinion and want to know everything that you like and dislike about certain programs.

Our goal is to serve you better with programming that *you want*. Your name was selected at random from a list of all residents living in Middletown. We promise that no one will come to your door or call you on the phone concerning this survey. Your answers will be *strictly confidential*. The final report will make no mention of the specific people that participated in this study.

Please complete the enclosed questionnaire as soon as possible. We have a deadline of Wednesday, February 12 for receiving your responses. The survey will probably take less than seven minutes for you to complete and place inside the prepaid return envelope. Again, I thank you for your time and consideration. I hope that you can help us out.

Sincerely,

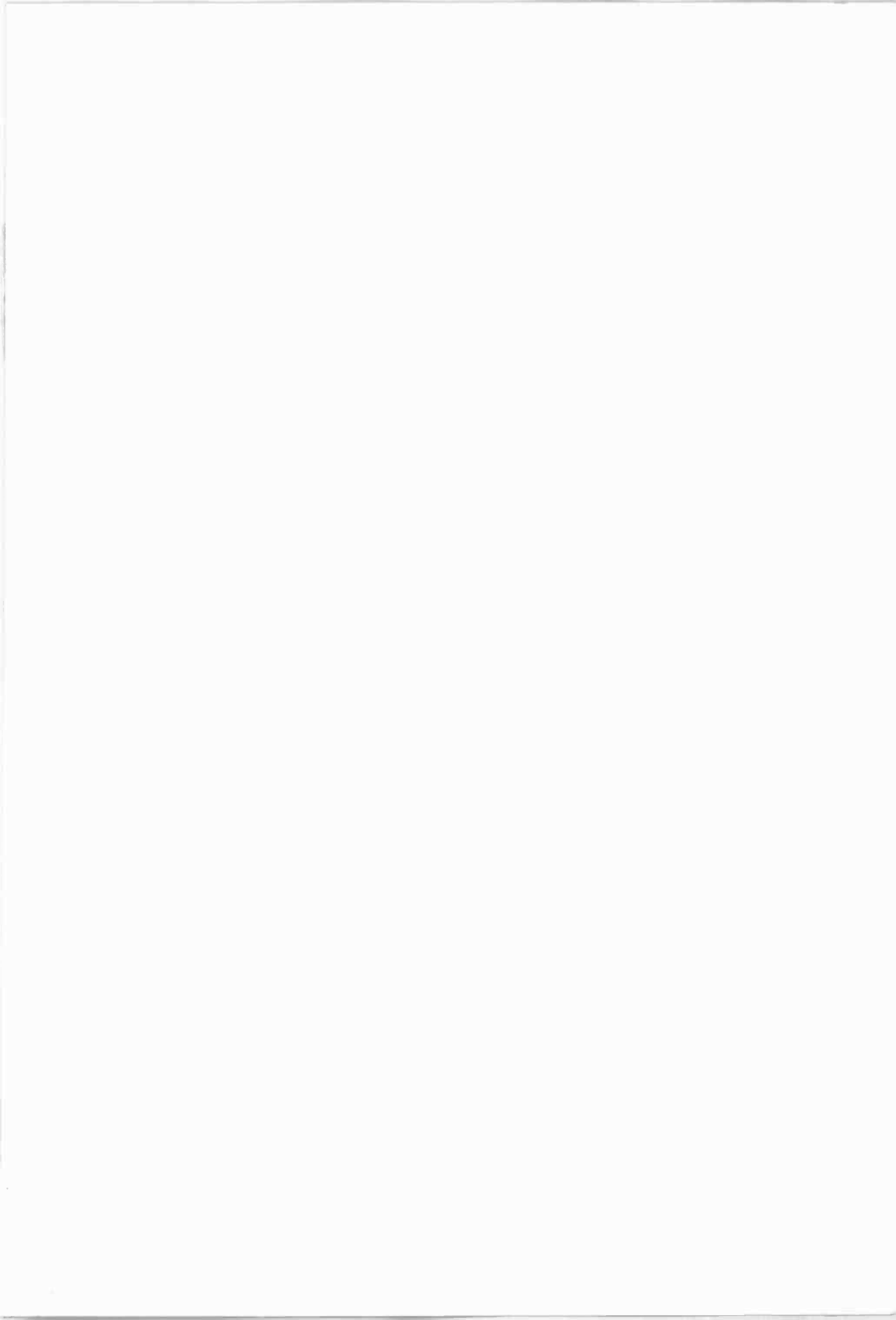
Frank Jones
President

About the Author

Walter McDowell, Ph.D.

Before entering academia, Walter McDowell spent over two decades in television broadcasting, including a dozen years as Director of Marketing and Creative Services for ABC affiliate WFTV, Orlando. While working his “day job” at WFTV, he taught evening classes for many years at the University of Central Florida. In 1994 he left the private sector and entered the Ph.D. program in Mass Communication at the University of Florida, Gainesville, where he continued teaching while earning his Doctorate. Upon completion of his advanced degree in 1997, he joined the faculty of Southern Illinois University at Carbondale, College of Mass Communication and Media Arts, where he currently teaches various courses in media management.

In 1999, McDowell, along with Alan Batten, President of AB Communications, wrote another NAB publication, *Branding TV: Principles and Practices*.



TROUBLESHOOTING AUDIENCE RESEARCH

DETECTING
PROBLEMS IN DESIGN,
EXECUTION AND
INTERPRETATION

Troubleshooting Audience Research is designed to help non-research media professionals make better use of the research they encounter on an everyday basis. It highlights the most important attributes of good (and bad) audience research in a clear and easy-to-understand way that de-mystifies the research process for both novices and experienced man-

agers alike. Its focus on practical applications is ideal for busy professionals who want to get the most out of audience research.

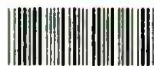
"Troubleshooting Audience Research provides readers with the knowledge they need to understand that research used in the wrong way can lead to very questionable conclusions, and counsels them on how to avoid common traps in making better research-based decisions."

Tom McClendon
Vice President & Director of Research
Cox Broadcasting

National Association of

NAB
BROADCASTERS

1771 N Street NW
Washington, DC 20036
To order: 1-800-368-5644
www.nab.org/nabstore
Item #3902



3902

Price: \$24.95

DESC. AUDIENCE RESEARCH

EACH

1-800-368-5644
06-11-03-568