This is to certify that the

dissertation entitled

THE EFFECT OF HUMOR IN INSTRUCTIONAL TEXT
ON LEARNING, INTEREST, AND ENJOYMENT:
IS GOOD HUMOR JUST FOR ICE CREAM?

presented by

Lorinda Marie Sheppard

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Counseling,
Educational Psychology,
and Special Education

Major professor

Date 6/21/02

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
FEB 1 0 2004 0 5 0 3 0 4		
MAR 2 4 2006 0 4 2 2 0 6		

THE EFFECT OF HUMOR IN INSTRUCTIONAL TEXT ON LEARNING, INTEREST, AND ENJOYMENT: IS GOOD HUMOR JUST FOR ICE CREAM?

By

Lorinda Marie Sheppard

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Educational Psychology, and Special Education

2002

ABSTRACT

THE EFFECT OF HUMOR IN INSTRUCTIONAL TEXT ON LEARNING, INTEREST, AND ENJOYMENT: IS GOOD HUMOR JUST FOR ICE CREAM?

By

Lorinda Marie Sheppard

This study examined the effect of humor in text-based instruction on learning, interest, and enjoyment. A review of the literature indicated both theoretical and empirical support for the potential effects of humor on learning, and research suggests that interest and enjoyment, in addition to being outcomes of humor, may also influence learning.

For this study, a humorous version and a nonhumorous version were developed for each of two chapters from a book on basic experimental psychology; one chapter was on how to get experimental ideas, and the other chapter was on how to be fair with science (i.e., research ethics). Subjects read the humorous version of one chapter and the nonhumorous version of the other chapter. After reading each chapter, subjects answered several questions regarding how interesting and relevant they found the chapter, how much they enjoyed the chapter, how much they believed they learned from the chapter, and how well the chapter held their attention. Subjects were also asked what they liked and disliked about each chapter, what helped them learn the information in each chapter, and what the two most important ideas were that they learned from each chapter. Subjects then completed a short-answer and multiple-choice quiz. After reading both chapters and answering questions specific to each chapter, subjects were asked to compare the two

chapters on several measures. Two weeks later, subjects completed another quiz for each chapter to determine long-term retention of the information.

On the tests of immediate learning, humor was associated with increased learning for the experimental ideas (EI) chapter, but not for the scientific fairness (SF) chapter. There was no difference in long-term retention as measured by the delayed tests of learning. Humor was associated with increased enjoyment, increased situational interest, and improved attention for both chapters. Humor had no effect on personal interest or perceived relevance for either chapter. Differences in the effect of humor on immediate measures of learning may be due to differences in the amount, type, and integration of humor in the two chapters, and to differences in the cognitive burden placed on readers by the content of the chapters.

In comparisons of the two chapters, subjects reported that the experimental ideas chapter was easier to stick with, more enjoyable, and preferred for leisure reading, regardless of use of humor. Subjects reported they would more likely to find the nonhumorous version in a textbook, regardless of topic.

Additional research is needed to determine the effects of humor in a classroom situation, and to determine if there is an optimal amount of humor in instructional text. Further experimental studies should explore the effect of humor on learning when the materials are more closely matched for type and amount of humor, general interest, and difficulty of content.

ACKNOWLEDGEMENTS

This dissertation never would have been completed without the support, encouragement, and contributions of a number of people.

A huge thank you goes to **Steve Yelon**, advisor for both my master's and doctoral programs. Without his phenomenal advising and mentoring, it's unlikely I would have begun the Ph.D. program, and even less likely I would have finished. Fortunately, Steve never really believed me when I said, "I just can't do this anymore." His response was always, "Let's just get you through this semester, and worry about the rest later."

Fran Yelon also deserves a thank you for her kind and encouraging words, especially during times when I was frustrated and disheartened. As Steve's last advisee, I hope my graduation will mark an end to the frantic and tearful phone calls she has had to field over the last many years from Steve's students.

Thank you to the members of my dissertation committee: **Jere Brophy**, **Punya Mishra**, and **Ed Wolfe**. Their willingness to work with me in a collegial and collaboratory manner was much different from the adversarial relationships some students have with their committees. They worked with me closely through all stages of this project so that the design, data collection, analysis, and interpretation went as flawlessly as possible.

Thanks to **David W. Martin**, author of *Doing Psychology Experiments*, the book that served as a source for the readings I used in this study, for allowing me to use his book and for corresponding with me about the use and potential benefits of humor in text.

None of this would have been possible without my parents, **Ike and Donna Sheppard**, who, while never placing any pressure on me to complete the program, were unwavering in their belief that I could eventually obtain a doctoral degree. Dad also provided a strong role model: the fact that he had a Ph.D. influenced me immeasurably in believing this was a worthy goal. Mom deserves special recognition for the emotional support she provided.

Thanks to friends **Donna Murrell**, **Anne Carrigan**, and **Telaina Eriksen** for serving as raters and creators of the humor used in the materials. They also made sure I remembered there was life outside of graduate school..

Thanks to friends **Dan Kurkowski** and **Kevin Music** for rating the humor used in the materials and providing needed breaks from the drudgery of schoolwork.

Thanks to fellow graduate students **Gina Cervetti, Michael Pardales, Mark Girod, Lisa Sensale**, and **Carolyn Jaynes**, who provided a support group during the difficult times, and who inspired me with their dedication to the pursuit of scholarly excellence. I am in awe of them.

I also appreciate **Lee Kokinakis, Debby Sleight**, and **Joe Brocato**, who were all graduate students when I met them and who preceded me in obtaining their doctorates, for the example they set and for their words of encouragement as I pursued my degree.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Statement of the Problem

Student motivation contributes to school performance: students who are motivated to learn are likely to learn more (Brophy, 1998) (see Figure 1-1). But how can we facilitate motivation to learn in traditionally less motivating situations, such as when students must engage independently with a textbook? Textbooks are frequently used as a supplement to traditional expository instruction, and often as a part of discovery learning as well. But students report they do not find academic reading to be interesting or enjoyable, and they are not affectively involved with such reading (Lorch, Lorch, & Klusewitz, 1993). Unfortunately, instruction will only be effective if students are engaged with it (Okey & Santiago, 1991). All text-based instruction has an affective component, and a "heavy" approach, such as that found in most textbooks or in a formal lecture, is less likely to draw learners in (Rowell & Van Kirk, 1978). How can we motivate students to want to read textbooks?

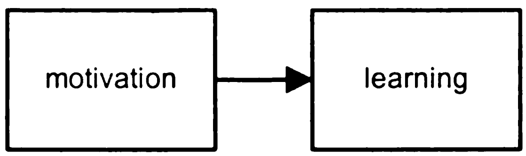

Figure 1-1. Motivated students are likely to learn more.

Past research on text-based instruction provides several ways to increase learner motivation, including increasing the perceived relevance of the instruction (Means, Jonassen, & Dwyer, 1997); using headings, overviews, and summaries (Lorch, Lorch, & Inman, 1993); activating prior knowledge (Spires & Donley, 1998); and increasing the concreteness or ease of imagery (Sadoski, Goetz, & Fritz, 1993). Research on classroom instruction suggests additional methods for increasing student engagement that may be useful in the development of text-based instruction, including instructor immediacy and use of humor.

Understanding the differences between academic reading and leisure reading may provide clues regarding how we can facilitate student engagement with texts. Lorch, Lorch, and Klusewitz (1993) found that students reported academic reading to be cognitively demanding, requiring memorization, self-testing of understanding, and close attention. Students also reported that they read academic material more slowly. In contrast, they reported leisure reading to be more enjoyable, more interesting, more emotionally involving, and less cognitively demanding. Students reported that they read leisure material more quickly. They were more likely to anticipate and to visualize when reading for leisure, and they were less critical and distractible. This was true even when the leisure reading was intellectually challenging or when the purpose was to learn something in an area of interest.

When we develop instructional text, how can we tap into the aspects of leisure reading that reduce students' cognitive burden and increase learner interest in the text? Perhaps there are techniques we can use to make the reading more enjoyable and interesting, such as by accessing students' positive feelings. Toward that end, Alexander, Kulikowich, and Jetton (1994) asserted that interest stimulates depth of processing which, in turn, enhances learning. Isen, Daubman, and Gorgoglione (1987) suggest that positive affect may enhance learner retention by increasing the connections the learner makes and thereby increasing the learner's integration of new information with prior knowledge.

Positive affect can be enhanced through the use of humor, and many opinion articles have been published on the perceived benefits of humor in education and training (e.g., Parrott, 1994; Wandersee, 1982; Korobkin, 1988; Peterson, 1980; Robbins, 1994). Controlled studies of humor in the classroom (e.g., Kaplan & Pascoe, 1977; Bryant, Comisky, Crane, & Zillmann, 1980) provide support for the positive effects of humor on student recall. Perhaps humor incorporated into textbooks can increase learning, in addition to enhancing learner interest and enjoyment of the topic of the reading. Figure 1-2 shows the potential outcomes of humor incorporated into text-based instruction.

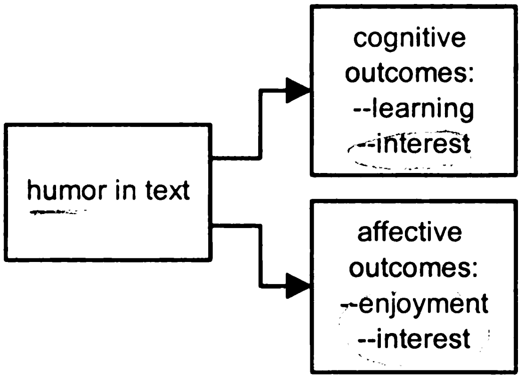

Figure 1-2. Humor's potential effects on both cognitive and affective outcomes, including learning, interest, and enjoyment

Figure 1-3 shows the hypothesized potential benefits of humor in text, including some of the intermediate variables as inferred from the literature.

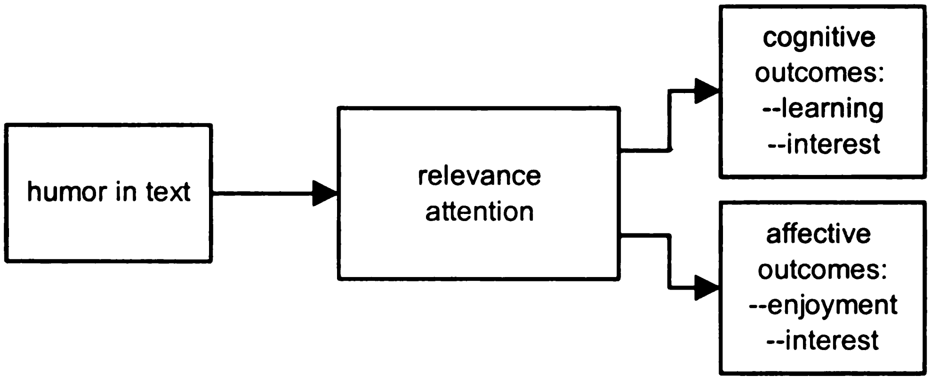

Figure 1-3. Mediating variables that may influence the effects of humor in

Briefly, these variables are defined as follows and will be measured as indicated. (See chapters 2 and 3 for more in-depth discussions of these variables, and see chapter 4 for the specific items associated with each scale.)

- **Humor** is a quality that makes something seem funny or amusing. Humor will be measured by self-reports of subjects' perceptions and, to a lesser extent, by evaluations of other readers who identify humorous items in the text.
- **Relevance** refers to whether the text meets the personal needs or goals of the learner. Perceived relevance will be measured by self-reports.
- **Attention** refers to mental concentration and will be measured by self-reports.
- **Learning** refers to retention and understanding of information and will be measured by recall and recognition quizzes. Subjects will also be asked to report on their perceptions of their learning.
- **Interest** is a feeling of intentness about something and can affect learners both cognitively and affectively, increasing both the learners' knowledge of the topic and their positive feelings (Krapp, Hidi, & Renninger, 1992). Interest can be specific to an individual or associated with the particular situation. Interest will be measured by self-reports.
- **Enjoyment** is the act or state of getting pleasure from something and will be measured by self-reports.

If we can find out what effects humor in text has on learning, interest, and enjoyment, then we will have clues that can help us address the much larger problem of student motivation in learning. Thus, the central question of this study is: Does humor in text-based instruction affect learning, interest, and enjoyment?

Rationale and Significance of the Study

The purpose of this study is to examine the effect of humor in text-based instruction on learning, interest, and enjoyment. Although the model presented earlier shows multiple potential effects of humor, I have chosen to focus on humor's effects on learning, interest, and enjoyment because I believe these are three significant goals of instruction. Obviously any techniques we can discover that will help improve learning (as measured through tests of recall and recognition) are desirable. However, enhancing learner interest in and enjoyment of a topic are also worthy goals that may in turn lead to further study.

When I tell other people that I am studying humor and learning, the most common response I get is an assertion that humor enhances learning, followed by an anecdote relating a humorous incident in a classroom or training session. However, despite most people's perceptions of the positive effect of humor on learning, the numerous books and articles written on how to incorporate humor into your classroom or training sessions (e.g., Loomans & Kolberg, 1993; Watson & Emerson, 1988), and the numerous studies on the positive effects of humor in the classroom (e.g., Ziv, 1988; Kaplan & Pascoe, 1977; Bryant et al., 1980), there are very few studies on the effect of humor in text on learning. In fact, Schmidt (1994) noted that no comparisons had been made between humorous and nonhumorous materials that were matched. He further stated that "in an appropriate test of the effects of humor, the content of the material must be controlled, ruling out explanations that are based on interest" (p. 953). A review of the literature has indicated no such comparisons. In this study, I intend to look at the effects of humorous vs. nonhumorous materials that are matched on content and explore humor's effects on learning, interest, and enjoyment.

Overview of the Study

To answer the central question— Does humor in instructional text affect learning, interest, and enjoyment?—I developed humorous and nonhumorous text materials and assessed reader learning via recall and recognition, as well as interest and enjoyment. Readings were adapted from *Doing Psychology Experiments*, a textbook on experimental psychology by David Martin. Two versions of two different sections of this text were developed—one version of each section was humorous, and one version had the humor removed but retained Martin's informal, conversational tone. Since a conversational tone has been found to increase reader perception of relevance and enhanced interest (Rowell & Van Kirk, 1978; Means et al., 1997; Paxton, 1997), this study explored the effect of humor on these variables over and above what might already occur as a result of the conversational tone. Subjects read the humorous version of one chapter and the nonhumorous version of the other chapter. After each reading, subjects were given two quizzes, one open-ended and the other multiple choice. To see if the independent variable was operating, I asked subjects to rate how humorous they found each reading. I also asked subjects how interesting, enjoyable, and relevant they found each reading, and how well each reading held their attention. Two weeks after participating in this part of the study, subjects completed another quiz, because other researchers and my own pilot study indicated that humor may enhance learner recall, but only for delayed recall and not for immediate recall.

There were significant differences in subjects' responses to the two chapters on several of the variables, so I was unable to collapse the data across topics as I had planned, and therefore unable to compare humorous readings to nonhumorous readings. For each of the chapters, I looked at the effect of humor on 1) learning, as indicated by measures of recall (measured by the open-ended quizzes) and recognition (measured by the multiple-choice quizzes) of facts and concepts in both immediate and delayed testing conditions, 2) interest, 3) enjoyment, 4) relevance), and 5) attention. I also had subjects compare the two chapters they had read on a variety of dimensions.

Humor in Text-Based Instruction

I chose to study humor in text materials instead of in the classroom because text allows for a greater amount of experimental control than is possible in a classroom situation. Studies suggest that spontaneous humor in the classroom is more effective (i.e., more humorous) than planned humor, which makes a controlled study of humor in the classroom difficult. Also, some instructors are naturally more humorous than others, which means they may have to suppress a natural tendency toward humor in order to provide a control condition. I am not convinced that such suppression affects only humor in the classroom without affecting other aspects of the instructor's style.

In addition, the lack of studies on the effect of humor in academic text makes this area ripe for study.

Overview of the Dissertation

This dissertation has five chapters:

Chapter 1: Introduction

As you have already seen, the introduction gives an overview of the research project, including a statement of the problem, the rationale and significance of the study, the purpose of the study, and an overview of the study.

Chapter 2: Literature Review

In order to establish a model of the theory I tested, Chapter 2 provides a review of relevant literature, particularly in regarding the interrelationships of humor, interest, enjoyment, and learning. This review also documents the parts of the model presented in chapter 3 that are supported by prior research.

Chapter 3: Methods

The methods chapter provides a detailed description of how the study was carried out, including descriptions of the subjects, instruments, and procedures.

Chapter 4: Results

Chapter 4 describes the analyses that were run on the data.

Chapter 5: Discussion

The discussion compares the hypothesized results to the actual results, and provides possible explanations for discrepancies.

CHAPTER 2: LITERATURE REVIEW

In this chapter, I provide evidence from the literature indicating that it is reasonable to expect humor in instructional text to affect learning, interest, and enjoyment. As stated in chapter 1, I have chosen to study humor's effect on learning, interest, and enjoyment because I believe these are important goals of instruction. Obviously we want our students to learn, but if we can stimulate their interest in learning and increase their enjoyment of learning, we may increase the likelihood that they will seek out, or at least be open to, additional experiences with the subject matter. Figure 2-1 illustrates the hypothesized effect of humor in instructional materials.

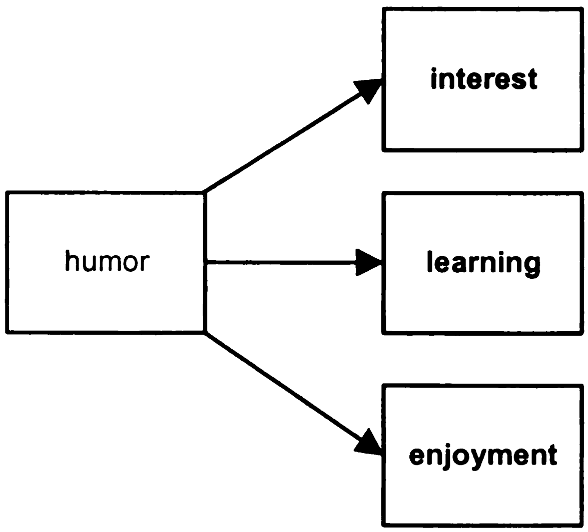

Figure 2-1. Hypothesized effects of humor in instructional

Student learning from text is influenced by numerous variables, including prior knowledge (Alexander, Kulikowich, & Schulze, 1994); engagement (Sadoski, Goetz, & Rodriguez, 2000); personal and situational interest (Krapp et al., 1992); affect (Bless, 2001; Fiedler, 2001); text structure, including use of headings, overviews, and summaries (Lorch, Lorch, & Inman, 1993); purpose of reading (Narvaez, van den Broek, & Ruiz, 1999); ease of imagery, or concreteness (Sadoski et al., 1993); and author visibility (Nolen, 1995; Paxton, 1997). Some of these variables were beyond the scope of this study and were either held constant across conditions (e.g., text structure) or assumed to wash out through random assignment of subjects to conditions (e.g., prior knowledge).

Students who are motivated to learn are likely to learn more than unmotivated students (Brophy, 1998). Motivation is strongly linked to interest and enjoyment: according to (Deci & Ryan, 1985), behavior that is intrinsically motivated is, by definition, interesting and enjoyable. I propose that humor in text, if it is found to have an effect on learning, influences learning in part by increasing reader enjoyment and interest. In other words, I hypothesize that interest and enjoyment not only are important outcomes of instruction by themselves, but also may influence learning. Figure 2-2 shows this relationship.

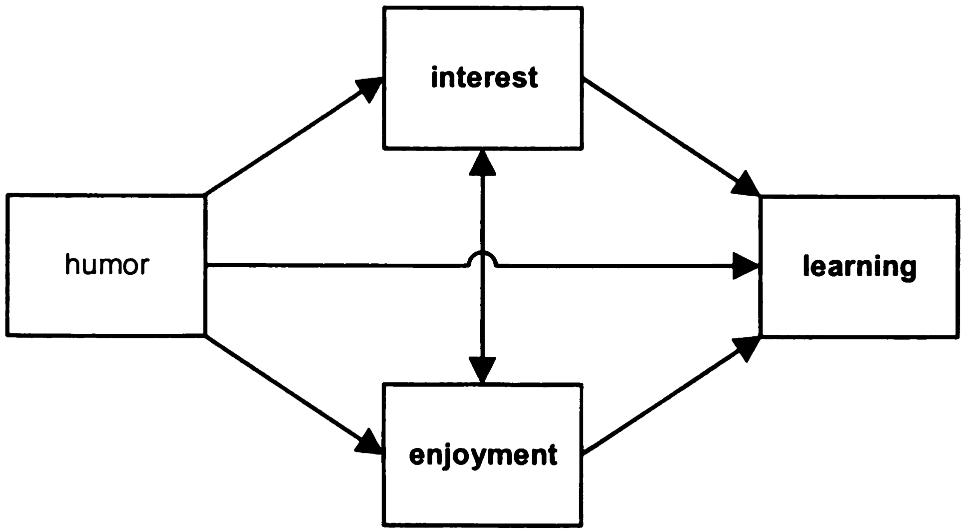

Figure 2-2. Interest and enjoyment as mediating variables between humor and learning

Humor may also aid learning by directing learners' attention and increasing perceived relevance, which facilitate encoding and retrieval of information. Attention and relevance may interact with interest and with each other, ultimately affecting learning. Figure 2-3 shows the model taking these variables into account.

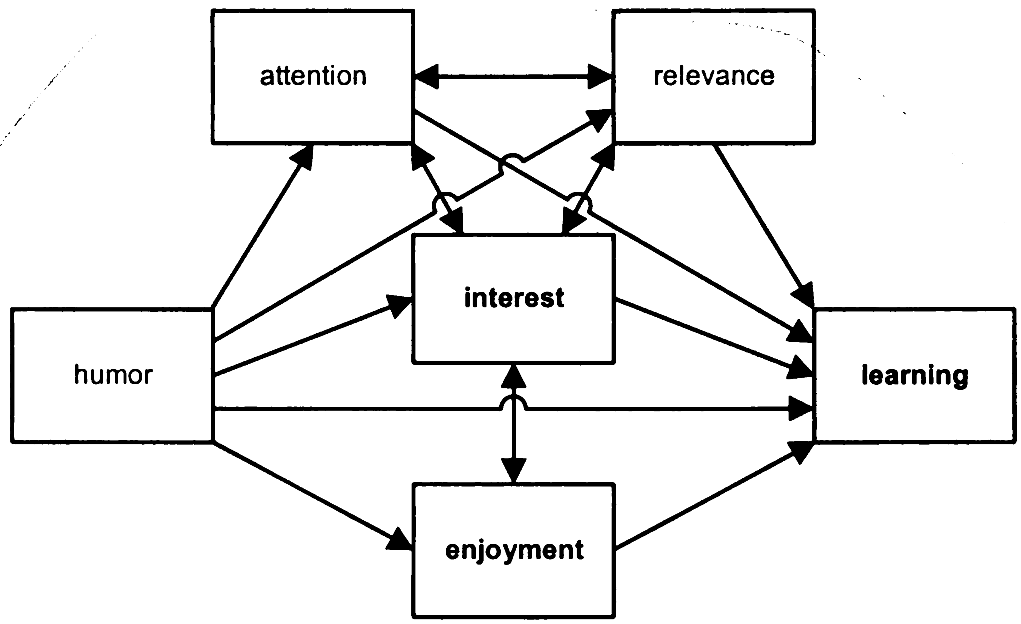

Figure 2-3. Hypothesized effects of humor in instructional text on both primary (i.e., learning, interest, and enjoyment) and secondary (i.e., attention and relevance) variables

In summary, the assertions upon which this study is based are as follows:

- Humor increases learning of the topic presented humorously.
- Humor increases interest in the topic presented humorously.
- Humor increases enjoyment of the topic presented humorously.
- Interest increases learning.
- Enjoyment increases learning.
- Interest and enjoyment influence each other.

- Humor increases attention, which interacts with interest to increase learning.
- Humor increases relevance, which interacts with interest and attention to increase learning.

Below, I begin my review of the literature by summarizing research related to the effect of humor on learning. I then discuss the expected effect of humor on my other two outcome variables, interest and enjoyment, and the effects of these variables on learning. I also discuss the mediating effects of relevance and attention on learning, and their relationship to interest.

The Nature of Humor

Humor is "the quality that makes something seem funny, amusing, or ludicrous" (Agnes, 2000). Any study on of the effects of humor should perhaps begin with a discussion of the nature of humor; specifically, how humor works. Numerous theories of humor have been proposed: in a 1998 article describing his theory of humor, Veatch noted that 619 papers on various theories of humor had been written to date. The main humor theories are the cognitive–perceptual or incongruity theory, the social–behavioral or superiority theory, the arousal theory, and the psychoanalytic theory (Johnson, 1990; Goldstein, 1993).

The cognitive–perceptual or incongruity theory proposes that humor is the result of a disparity between what is expected and what actually happens. Humor is found when ideas that do not belong together are joined. This theory posits that humor is a two-stage process: the audience first has to perceive the disparity and then resolve or understand it (in other words, they have to "get" the joke). Perception of the disparity is subjective and may be affected by the "values, attitudes, knowledge, and expectations" of the audience (Goldstein, 1993, p. 247).

The social–behavioral or superiority theory suggests that humor is the result of disparagement of others. This theory is based on the idea that we laugh at the failings and mistakes of others so that we can feel superior to them (Goldstein, 1993). Barreca (1991) noted that men and women use this type of humor differently: men tend to disparage those they consider inferior, such as people with lower intelligence or those with physical handicaps, while women will make fun of those in positions of power, such as supervisors and bosses.

The arousal theory sees humor as a release of tension. According to this theory, tension increases as the punch line nears. When the punch line is delivered, arousal is increased sufficiently to produce amusement or laughter (Goldstein, 1993).

The psychoanalytic theory proposes that "laughter [can] be symptomatic of conflict, repressed sexuality, and anger," and that humor is a powerful defense mechanism (Goldstein, 1993, p. 247). Freud noted that "tendentious" jokes are sexual or aggressive in nature and so tend to elicit a stronger emotional reaction than "innocent" humor (Veach, 1998).

In his (unnamed) theory of humor, Veach proposed that "humor occurs when it seems that things are normal while at the same time something seems wrong" (p. 164). In other words, the situation is seen as predominantly normal, while at the same time a violation of the way things ought to be is perceived. Veach argues that this theory is more accurate than the more popular and widely cited incongruity theory on the grounds that many incongruities do not necessarily produce humor (note, for example, the many incongruities that arise during scientific study that do not elicit a humorous reaction).

Regardless of the type of humor or the theory used to define it, humor is largely a function of the perceiver. Certain events may strike some people as enormously funny, yet fail to elicit even a smile in others. Despite this, a fair amount of humor will be universally rated as humorous or at least recognized as an attempt at humor.

Effects of Humor on Learning

Despite individual differences in the perception of humor, there are some types of humor that are more universally accepted than others. In an attempt to verify that the intervention intended to be humorous was perceived by subjects as such, I asked subjects in this study to rate the materials on their "humorousness."

There is theoretical support for the effect of humor on learning. According to dual coding theory, comprehension and memory of text are affected by both verbal (e.g., text, speech) and nonverbal (e.g., imagery, affect) systems (Sadoski et al., 1993). The effects of these two systems are additive—we understand and remember more if both of these systems are engaged than if we rely on only one. Obviously, text will engage the verbal system. Humor, on the other hand, will enhance comprehension and memory by stimulating affect. Humor and text combined should have a powerful effect on learning.

Studies of Humor and Memory

Humor has been proposed to have numerous effects that increase memory for humorous information. Some researchers (e.g., Karis, Fabiani, & Donchin, 1984; Schmidt, 1991) argued that the incongruity component of humor (i.e., that humor results when ideas are presented out of context or in unusual combinations) leads to an increase in attention. Alternatively, Schmidt and Williams (2001) hypothesized that the attempt to resolve the incongruities in order to understand the humor involves greater cognitive processing, which leads to an increase in learning. Yet another hypothesis is that subjects rehearse humorous material more, which leads to an increase in learning and memory of the humorous items (Schmidt, 1994).

A study by Schmidt and Williams (2001) attempted to separate the effects of humor from those of incongruity and "bizarreness." They found that humor, separate from incongruity and bizarreness, enhanced recall. They concluded that humor itself is the "potent mnemonic variable" (p. 307). The actual resolution of the humor or the perception of humor that resulted, rather than an *attempt* to resolve the humor, enhanced recall. Humor enhanced memory even when the influence of intentional verbal rehearsal was minimized. Schmidt and Williams (2001) concluded that their findings weaken the argument that rehearsal explains the effect of humor on memory.

In an additional study, Schmidt and Williams (2001) found that the "humorousness" of cartoons was directly related to subjects' memory of them: the most humorous cartoons were remembered better, and the cartoons rated in the upper half of "humorousness" were recalled better than those in the lower half. These results indicated that humor itself, rather than incongruity, resolution of incongruity, bizarreness, or distinctiveness, enhances recall for cartoons (Schmidt & Williams, 2001).

Studies of Humor in the Classroom

There is moderate support for the effect of humor on learning in the classroom. Although publications on humor and adult learning appear to be largely anecdotal (e.g., Parrott, 1994; Wandersee, 1982; Korobkin, 1988; Peterson, 1980; Robbins, 1994), a few controlled studies of the use of humor in the classroom have appeared in the literature (e.g., Bryant et al., 1980; Ziv, 1988). Perhaps one reason for a lack of empirical research on humor in the classroom is its almost inherent uncontrollability: doing a controlled study either would require instructors who are naturally humorous to suppress this tendency, perhaps resulting in differences in the way they interact with their classes more globally, or would require instructors who are not humorous to attempt to be funny on demand. Since the "best" classroom humor is perceived as spontaneous (Bryant et al., 1980), planned humor by nonhumorous instructors would probably yield unnatural, and possibly uninterpretable, results.

Studies of Humor in Text

Perhaps even more surprising than the lack of controlled studies of humor in the classroom is the almost complete lack of studies of humor in instructional materials. After all, there is a large amount of humor available in printed form (for example, in books, magazines, and newspapers). Unfortunately, humor does not appear to be prevalent in instructional materials. As I was preparing for this study, I sent an e-mail to 61 people, including family members, friends, professors, fellow students, and current and former coworkers, requesting references to humorous instructional materials they knew of. I received about a half dozen recommendations, none of which were suitable for my study (e.g., they were cartoon-based rather than text-based, or the author's writing style was conversational rather than truly humorous). Humor appears to be a casualty of "textbookese," the third-person expository style used in most textbooks that makes few, if any, references to either the reader or the writer (Crismore, 1984). Rather, most textbooks give the impression that their contents are delivered from an omniscient "other" who lacks any apparent viewpoint, opinions, biases, or personality. A sense of humor would certainly be out of place in this style of writing.

Despite the dearth of humor in textbooks, there is evidence to suggest that humor in instructional materials may contribute to both learning and the perception of learning. In the evaluation of a basic computer literacy self-study module, 67 percent of respondents felt that the humor contained within the module had helped them learn (Sheppard, 1998). This study is supported by a classroom study conducted by Mogavero (1979), in which 95 percent of students felt that humor increased the effectiveness of instruction. In addition, Zillmann and Bryant (1983, cited in Robbins, 1994) found that humor that is well integrated with the material can lead to greater retention of the information and make learning more enjoyable.

In 1994, Schmidt noted that no comparisons had been made between humorous and nonhumorous materials that were matched. He further stated that "in an appropriate test of the effects of humor, the content of the material must be controlled, ruling out explanations that are based on interest" (p. 953). A review of the literature since that time has revealed no studies that satisfy these criteria. In this study, I intend to fill that gap by looking at the effects of humorous vs. nonhumorous materials that are matched on content and exploring humor's implications for learning, interest, and enjoyment.

Effects of Humor on Enjoyment

Although ideally humor will enhance learning, another important outcome of humor in instructional materials is to increase learner enjoyment. **Enjoyment** is a specific form of positive affect that refers to pleasurable experiences. **Affect** is a broad term that refers to both mood (an enduring trait of an individual that is not tied to any definite event) and emotion (a more intense, short-lived state that generally has a specific cause) (Forgas, 1991). A learner who enjoys the learning experience may be more receptive to, and perhaps even seek out, additional experiences with the topic.

There is evidence in the literature to support a link between humor and enjoyment. Humor is enormously popular in our society. Comedy is among the most popular television genres, and as of the early 1990s, over 40 percent of the all-time top 100 rental movies and the all-time highest-rated television shows in the United States were comedies (Goldstein, 1993). Humor facilitates a change from a bad or mediocre mood to a more positive mood (Zillmann & Bryant, 1985; Zillmann, 1988; Bryant & Zillmann, 1989). Comedy allows individuals to "regulate their own emotional states"—people may actively seek out humor and comedy in order to positively affect their mood (Zillmann & Bryant, 1985). This study will determine if the link between humor and enjoyment holds true for humor in instructional materials.

Effects of Humor on Interest and Attention

Interest has been defined as "liking and willful engagement in a cognitive activity" (Schraw & Lehman, 2001, p. 23). Interest is a feeling of intentness about something that can affect learners both cognitively and affectively, increasing both the learners' knowledge of the topic and their positive feelings (Krapp et al., 1992). **Attention** refers to capturing the learners' interest and stimulating curiosity to learn (Keller, 1987); by definition, then, attention is closely related to interest. Interest is demonstrated through behaviors that include attention (Krapp et al., 1992), and interest results in spontaneous rather than selective allocation of attention (Hidi, 1990).

Potential benefits of humor in the classroom include increased interest and attention, increased comprehension and retention of the material, increased motivation toward learning, increased satisfaction with learning, more positive attitudes, and decreased boredom (Korobkin, 1988; Parrott, 1994); some of these benefits may also apply to print materials. In addition, the use of humor in the classroom enhances the student-teacher relationship, particularly students' perceptions of instructor immediacy (i.e., physical and psychological closeness) (Gorham & Christophel, 1990). Based on the studies of visible authors (defined as authors who write in the first person and reveal personal opinions and other aspects of themselves), it may be reasonable to believe that humor in text can also take advantage of some of the benefits of instructor immediacy.

Despite our love of humor and its apparent links to learning, if there is too much humor in instruction, the presenter may cross the line between having the purpose of his or her presentation interpreted as informative and having it interpreted as entertainment. But humor in appropriate amounts may help maintain student interest and attention, thereby resulting in greater enjoyment of learning. Support for this assertion comes from Ziv (1988) and Bryant et al. (1980), who found that three to four instances of humor during an hour classroom lecture was not only the average (Bryant et al., 1980), but also the ideal (Ziv, 1988). Ziv also found that the humor must be related to the concepts presented in the class in order to enhance understanding; if the humor is unrelated to the material presented, it can actually distract students from the main ideas.

Effect of Humor on Perceived Relevance

Relevance refers to meeting the personal needs and goals of the learner (Keller, 1987). There is no evidence in the literature to link humor to perceived relevance.

However, there are certain devices—such as immediacy and visible author—that have been shown to increase learners' perceived relevance of material, and I am curious about whether humor is able to extend or compound these effects. Humor may also have a community-building mechanism that helps to increase perceived relevance.

Use of the word "you" has been shown to increase readers' perceived relevance of written materials (Means et al., 1997). Although use of a visible author has not been linked directly to increased perception of relevance, a visible author creates an atmosphere of collaboration with the reader (Paxton, 1997). Readers who are engaged by an author may be more likely to perceive the material as relevant—readers may unconsciously associate this engagement with the author as personal relevance. Humor goes beyond both the use of "you" and use of a visible author by endowing the author not just with visibility, but also with personality.

In addition, if humor helps to build a community, as argued by Cohen (1999), and the reader feels a part of this community, he or she may perceive an increased relevance of the materials. In other words, if readers feel that the humor is targeted toward them, they may be more likely to conclude that the instructional material in which the humor is embedded also is targeted toward them more directly.

Effects of Enjoyment on Learning

In addition to being an important outcome of instruction, learner enjoyment may also facilitate learning. Students' positive feelings toward a task or topic can increase their learning, both directly by fostering connections with prior learning and by facilitating use of general knowledge structures, and indirectly by increasing the likelihood that students will seek additional experiences with the task or topic.

Organization

Isen et al. (1987) hypothesized that "positive affect results in an organization of cognitive material, wherein either more or broader (or more integrated) categories are primed and utilized than is the case under control conditions. That is, under conditions of positive affect more different kinds of material are brought to mind and dealt with at the same time, there is a tendency to see relationships and group things together, and ideas that ordinarily would not seem to bear upon one another are seen as related or similar and holding implications for one another" (pp. 147-148). Building on what the student already knows and helping him or her form relationships between new knowledge and previous learning and experiences is one of the crucial elements of sound instruction.

Even minor differences in learners' affective states may have a substantial impact on their cognitive processes (Bless, 2001). In general, individuals experiencing a positive affective state are more likely to rely on general knowledge structures—that is, heuristics, stereotypes, or scripts they use on a regular basis. In other words, "individuals in a positive affective state may feel more confident about relying on activated general knowledge structures that are potentially applicable to the situation" (p. 11).

In contrast, individuals interpret a negative affective state as signaling a problem. When a problem exists, individuals may find it risky to rely on general knowledge structures, so they may instead attempt to deal with the problem by devoting their attention to the specific details. If we assume that problematic situations are deviations from normal situations, then relying on the knowledge we would usually apply during normal situations would probably not be the best approach.

Nonproblematic situations require less attention to the specifics. In these cases, the individual can rely on general knowledge structures, which saves processing resources. These spared resources can then be allocated to other aspects of the situation or to other tasks. This "transfer" of resources is highly adaptive, since it frees up attention for those aspects of the situation for which additional processing is potentially more beneficial and effective (e.g., inconsistent information). Individuals experiencing positive affect rely on general knowledge structures, which frees up additional attention resources to exploring the match between the knowledge structure and the available information, with inconsistencies triggering additional processing. Therefore, this mechanism allows for a flexibility in attention to the specifics of a situation.

General knowledge structures enable the individual to enrich the stimulus information by making inferences beyond the information given (Bruner, 1957). Individuals who go beyond the information given may sometimes come up with new, creative inferences (Bless, 2001). New situations confer a certain amount of risk, so it would seem highly adaptive to rely on general knowledge structures as a basis for making inferences, especially if the situation is safe rather than problematic (Bless, 2001). By extension, if we can trigger positive affect (e.g., enjoyment) through the use of humor in instruction, we may enable learners to activate general knowledge structures and make connections to prior knowledge.

Learning Sets

According to Fiedler (2001), "Over a long period of (ontogenetic or even phylogenetic) learning, organisms acquire different generalized learning sets for positive versus negative situations that emphasize either exploration and curiosity or avoidance and vigilance" (p. 93). In other words, a learning environment that promotes positive affect encourages exploration and creativity, while aversive learning situations (i.e., those that promote negative affect) maximize behaviors directed toward avoiding mistakes. In order to avoid the negative consequences of aversive learning situations, avoidance behavior must be reliable and attentive to potentially threatening stimuli. In contrast, in positive learning situations, the individual does not have to monitor any stimuli as carefully and can instead trust in internalized knowledge and behavioral routines. Negative mood (and the implied aversive learning situation) supports sticking to the stimulus facts and avoiding mistakes, while positive mood enables creative functions such as active generation and enrichment of the stimulus input with inferences based on prior knowledge. "Positive affective states signal the absence of dangers and obstacles so that the individual can take the risk of unusual, creative, exploratory behaviors" (Fiedler, 2001, p. 92). Even superficial and transitory affective cues, such as those manipulated by aspects of the situation, may be as effective in bringing forth certain learning sets (i.e., those associated with either positive or negative affective states) as more "genuine" emotions (Fiedler, 2001). I propose that the use of humor in text may provide affective cues that help elicit learning sets associated with positive affect.

As educators, it behooves us to use whatever means we have available to help learners make meaningful connections between what they already know and what they are learning. Isen et al. (1987) noted that "positive affect leads to a change in the way material is categorized, and this supports the more general formulation … regarding the influence of positive affect on cognitive organization" (p. 151). Oddly, however, Isen et al. obtained these effects after inducing positive affect by giving the subjects a small gift or juice and cookies, or by having subjects watch five minutes of a comedy video—in other words, *through methods that were unrelated to the topic under study.* This is consistent with Fiedler's (2001) assertion that superficial cues may be effective in stimulating the learning sets associated with positive emotions. But if favorable changes in cognition are observed even when the methods of improving affect are not related to the topic under study, what is the potential impact when affect is positively influenced via the instruction itself? This brings us back to the potential of humor that is integrated into the main ideas being learned (Ziv, 1988). Perhaps by integrating humor into texts, we can take advantage of the effects of positive affect on learning compounded, perhaps, by direct relatedness to the topic being studied.

Controlled studies offer support for the role of affect in learning. In a study of reader recall of feature articles from popular magazines, Sadoski and Quast (1990) asked subjects about factors that influenced their recall. They found that affect was the most frequently mentioned subjective characteristic—readers mentioned the presence of feelings as reasons for remembering text and lack of feelings as a reason for not remembering. The article that readers had, overall, rated the lowest in terms of affect was also the article that was most often reported as completely forgotten. This is consistent with a study of non-directed (i.e., incidental) learning from text, in which enjoyment of the text was positively associated with both perceived comprehension and text recall (Ryan, Connell, & Plant, 1990).

Visible Author

Rowell and Van Kirk (1978) asserted that self-instruction modules, by their very presence, have an affective component: learners form opinions about the subject based on the reading, and it is necessary for module designers to include "fun parts" to maintain the learners' attention and encourage students to stick with reading that may not appear interesting at the outset. They argued that a "predominantly cognitive, academic module" (p. 13) does not have to be dense and boring. They proposed instead that a conversational tone could help indicate to the student that the module presenter (writer, speaker, etc.) is aware of him or her as an individual, rather than just directing the information into a void. An author who is able to incorporate humor into instructional materials may also convey his or her enjoyment of the topic. Many students can recall an experience in which they found an instructor's enthusiasm contagious.

Along similar lines, recent research indicates that a visible author can have a positive effect on learning. Paxton (1997) had students think aloud as they read history texts that had either an invisible author (defined as an author who writes in the third person and reveals little or nothing of his or her own outlook or views) or a visible author. Students who read the text with the visible author engaged in mental conversations with the author, which led to a closer relationship with the information contained in the text. Students who read a text with an invisible author failed to engage with the text—Paxton (1997) wrote, "Student response to anonymous text might be characterized as passive boredom. These students showed practically no signs of engagement" (p. 245). In think-alouds while reading the anonymous text, "students gave little independent thought either to the author or to the subject matter. In short, the anonymous text imparted to students no feel for the subject of history, aside from the feeling that they would just as soon be doing something else" (p. 245). Readers of the visible text were "10 times more likely to give personal thought to the historical information presented in the text. ... The visible author sought to create a rhetorical atmosphere of empathy and collaboration with the reading audience. These students—even the one who disdained the visible author—reacted positively to this approach" (p. 245).

Similarly, Nolen (1995) found that women reading text containing a visible author interacted with the author while reading, which seemed to "influence the relations among comprehension, motivation, and affective response" (p. 47).

In a similar vein, Bergin (1999) noted that the way students interact with a text could be considered an aspect of social interaction. He further noted that the extent to which students perceive that an activity allows them to socialize (even if it's with the author of a text), they will tend to be interested in the activity. A text that contains humor may additionally increase learning by further reducing the social distance between the reader and the text.

In summary, enjoyment of learning may have multiple positive effects on learners, including facilitating the integration of new learning into prior knowledge, engaging learning sets, and enhancing student engagement with the text.

Effects of Interest on Learning

According to Krapp et al. (1992), interest emerges from an individual's interactions with his or her surroundings. They note that different researchers attribute different amounts of significance to each component in this person–environment interaction. Personal (or individual) interests tend to be relatively stable within an individual and are usually associated with increased knowledge and positive emotions. Personal interests are dependent on the individual's characteristics, attitudes, and orientations (Krapp et al., 1992); personal interest is enduring and context-general (Schraw & Lehman, 2001). An example from my own life is my interest in forensic science—I enjoy reading books by forensic psychologists, forensic anthropologists, and even forensic entomologists, and I try not to miss "Medical Detectives" and "Secrets of Forensic Science" on The Learning Channel. This is a topic I enjoy learning more about, and as a result of the time and effort I spend pursuing this interest, I imagine that I know more than the average person about certain aspects of forensic science.

In contrast, situational interests can be elicited more universally (i.e., in many individuals regardless of their individual interests) through the use of certain devices (e.g., narrative style) or characteristics (e.g., novelty) of the presentation (Krapp et al., 1992); situational interest is generally spontaneous and context-specific (Schraw & Lehman, 2001). Situational interest, which can be evoked by something in the environment, usually has a short-term effect. An example of situational interest from my life has to do with a Nova episode on automotive safety technology. Nova was the least objectionable show I found while channel surfing during a late lunch on a Friday afternoon. I had no interest in automotive safety, but the Nova developers have a remarkable ability to take a subject that I think of as a real snoozer and turning it into something fascinating. I was very involved during the episode, and I feel that I learned a lot about the evolution of automotive safety. I am probably more receptive to additional information on automotive safety, but I would not necessarily go out of my way to find out more about it.

Despite the distinctions made between individual and situational interest, these two forms of interest interact with and influence each other (Hidi, 2000). Specifically, situational interest may serve as the basis for the emergence of individual interests. In addition, personal interest may influence situational interest by moderating the effects of environmental variables on interest (e.g., an individual who has a personal interest in a particular topic may be bored by a book or article on that topic if it tells him or her nothing new). Krapp et al. (1992) refer to "actualized interest," which is demonstrated through particular behaviors, such as relatively effortless attention that is focused, prolonged, and accompanied by feelings of enjoyment. This actualized interest is the result of an interaction between internal and external conditions (i.e., individual and situational interest, respectively).

Schraw and Lehman (2001) described a taxonomy of interest. They divided individual interest into latent interest (i.e., "a long-term orientation toward a particular topic" [p. 29] and actualized interest (i.e., "a topic-specific motivational state that determines one's engagement style in an activity" [p. 30]. Latent interest, in turn, comprises feeling-related and value-related interest. In other words, latent interest is composed of things we are interested in because we like them or because we feel they have some value to us.

Situational interest, according to Schraw and Lehman (2001), is divided into text-based interest (i.e., interest based on properties of the information to be learned), task-based interest (i.e., instructions that change the individual's goals), and knowledge-based interest (i.e., the effect of prior learning on interest). By incorporating humor into text, this study will tie into aspects of both personal interest (specifically, the feeling-related interest that Schraw and Lehman described) and situational interest (specifically, text-based interest).

The presence or absence of interest, whether individual or situational, influences the type of learning strategies individuals employ. When learners are interested in what they are learning, they tend to use deep-processing strategies—they attempt to examine the material from different perspectives, make diverse connections, and identify and solve problems on their own (Krapp et al., 1992). In contrast, learners who are not interested use surface-level strategies, relying largely on memorization to "learn" material. Alexander, Kulikowich, and Jetton (1994) also noted, "Interest, particularly one's personal investment in the topic or domain, stimulates depth of processing in the content and, thus, enhances subject-matter learning" (p. 217). According to Hidi (1990), "Interest elicits spontaneous, rather than conscious, selective allocation of attention" (p. 549) and, therefore, may result in learning that requires less cognitive capacity and effort.

There is substantial support in the literature for the assertion that interest, whether individual or situational, influences learning. For example, Wade and Adams (1990) found that interest had "a powerful effect on recall for both good and poorer readers" (p. 331) and that seductive details and main ideas (both rated as interesting) were best remembered, while details supporting the main ideas were least well remembered. High interest is associated with greater recall of text information for both structured and unstructured recall tasks (Garner & Gillingham, 1991), and learner interest is significantly related to recall of main ideas, elaborations, and idea units (Schiefele & Krapp, 1996).

In summary, interest facilitates learning by directing learners' attention to the topic at hand and engaging deep-processing strategies which facilitate encoding and retrieval of information.

Effects of Attention on Learning

If learners don't pay attention to the instruction, it seems unlikely that they will get anything out of it. The idea behind Keller's "attention" is that grabbing and holding the students' attention is an essential characteristic associated with motivation, and is necessary for any learning to occur. Humor would fall into this component of the ARCS model, since it would be a part of the supporting strategy of "including emotional content" under attention's subcategory of perceptual arousal.

There are several theories about the relationship between interest and attention and their effects on learning. One theory suggests that increased interest leads to increased attention (Anderson, 1982, cited in McDaniel, Waddill, Finstad, & Bourg, 2000). A second theory is that interest results in a spontaneous allocation of attention, which frees up cognitive resources which can then be used for more flexible and efficient processing of the text (Hidi, 1990). A third theory agrees that interest results in spontaneous allocation of attention, but suggests that this enables the reader to focus on "organizational and situational elements implied by the text" (McDaniel et al., 2000) (p. 500). In contrast, readers who are not interested in what they are reading must expend resources to keep attention focused on merely extracting meaning from the text. Support for the effect of interest on attention comes from a study which found that interest did not affect recall of stories, but high-interest stories required fewer attentional resources for understanding than did low-interest stories (McDaniel et al., 2000).

Keller's ARCS model of motivation delineates four requirements that must be met in order for people to be motivated to learn: the student's **a**ttention must be gained and sustained, the material must be **r**elevant to the student, the student must be **c**onfident that he or she can learn the material, and the student must be satisfied with the process of the experience (Keller, 1987). One of the theories on which this model is based is the expectancy x value model of motivation. This expectancy x value model maintains that "the efforts that people are willing to expend on a task is the product of (1) the degree to which they expect to be able to perform the task successfully if they apply themselves (and thus the degree to which they expect to get whatever rewards that successful task performance will bring), and (2) the degree to which they value those rewards as well as the opportunity to engage in the processes involved in performing the task itself" (Brophy, 1998), p. 14). The value dimension of the expectancy x value model is most closely related to Keller's attention and relevance components, while the expectancy dimension is more closely linked to confidence and satisfaction (Keller, 1979).

Okey and Santiago (1991) support Keller's assertions, noting that regardless of how well designed instruction is, it will only be effective if learners are engaged with it. They further point out that motivational design and instructional design are complementary, rather than conflicting, ideas, and that instructional designers who wish to produce engaging instruction should link the ideas of motivational design to a theory of instructional design. Integrating humor into text may be one way to do this.

Although this study is not intended to determine the relationship between attention and interest, I believe that these are separate constructs and that humor may influence both. In my analysis of the data resulting from this study, I will explore the relationship between these two constructs. Such a close relationship between interest and attention will make it difficult to tease apart these two constructs. However, in my research, I will develop separate scales for individual interest, situational interest, and attention to determine if they can be measured separately.

Effects of Perceived Relevance on Learning

As stated earlier, relevance is one of the four elements of Keller's ARCS model necessary for individuals to be motivated to learn. In order for a student to sustain motivation, he or she must perceive that the instruction meets important personal needs (Keller, 1983).

Perceived relevance of material may influence the amount of mental effort learners are willing to expend to learn (Cennamo & Braunlich, 1996). In a study of print-based text that had been motivationally enhanced to increase perceived relevance, subjects who received the motivationally enhanced lesson remembered more information than those who received the nonenhanced version. The authors suggest that learners may take time to elaborate on the content of an enhanced lesson—specifically, they may stop to reflect on how the material applies to them and contexts in which they will be able to use the information.

Main (1992) notes that "the very act of establishing an instructional goal implies some value to the person, organization or society in its achievement" (p. 18). We enhance a student's desire to learn and to spend effort learning by helping them see the relevance to their job specifically or to their life more generally. Main further notes that the affective domain (including relevance) is particularly important to address in instruction that removes the interaction between the teacher and the student from the lesson delivery, such as occurs in text-based instruction.

A study by Means et al. (1997) found that perceived relevance was related to higher motivation and better posttest scores. Materials that had embedded strategies to enhance relevance increased learner motivation and achievement.

In summary, there is both theoretical and empirical support for the effect of relevance on learning, and subjects' perceived relevance can be enhanced in text through use of certain strategies.

Conclusion

In conclusion, there appears to be support in the literature for the links among humor, interest, enjoyment, and learning shown in Figure 2-4. Classroom studies provide

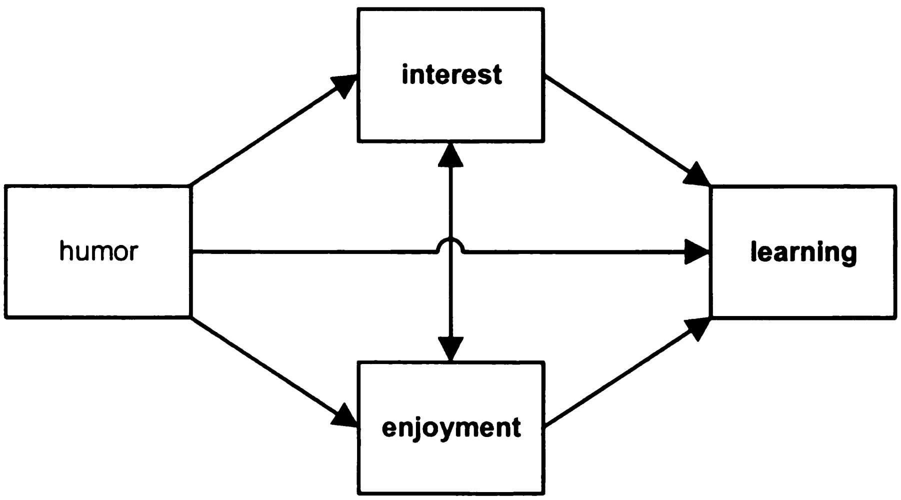

Figure 2-4. Hypothesized effects of humor, and relationships among the three main outcome variables

evidence for an effect of humor on learning. There are strong indications in the literature for the effect of interest on learning, and moderate indications for the effect of enjoyment on learning. Both interest and enjoyment are considered to be positive affective states and are linked in the literature on learning. Perhaps the weakest links, at least empirically, are those between humor and interest, and between humor and enjoyment; however, the fact that people actively seek out humor (in some cases because they want to change a poor mood into a good mood) implies that humor generates or enhances positiye affect (of which interest and enjoyment are subcategories). In theory, then, humor in text should enhance learning through its effects on interest and enjoyment. The research I propose will determine if this theory holds up in a controlled study.

CHAPTER 3: METHODOLOGY

Research Questions and Overview

My research question was: Does humor in instructional text affect learning, interest, and enjoyment? In this chapter, I describe the study designed to answer this question, including the participants, the materials, and the procedures. I also list the questions that guided the research. Finally, I describe limitations of the study.

Pilot Study

I conducted a pilot study on the effect of humor in instructional materials in the spring of 2000 to see if there was any evidence for humor's effects on learning, interest, and enjoyment. I recruited subjects through TE150, an introductory course in teacher education, in large part because these were the students I had access to—TAs for these courses allowed me into their classrooms to recruit subjects.

The research had two phases. During the first phase, 18 subjects provided demographic information and their initial reactions regarding how interesting and relevant they believed each of the two topics of the readings (i.e., getting an experimental idea and making instruction meaningful) was to them. Subjects read a humorous version of one chapter and a nonhumorous version of the other chapter. Readings were counterbalanced, with some subjects reading a humorous chapter first, and others reading a nonhumorous chapter first. Topics were also counterbalanced, with some subjects reading the chapter on getting experimental ideas first, and others reading the chapter on making instruction meaningful first. After reading each chapter, subjects answered both short-answer and multiple-choice questions about ideas contained in the text. Subjects also answered questions regarding the presentation of the material and their reaction to it, including how interesting, informative, humorous, and relevant they found each chapter. Subjects answered questions specifically about the acceptability of the humorous version and the effect of humor on the information being conveyed. Subjects compared the two versions on several measures, such as which they found easier to stick with, which they would be more likely to find in a textbook, and which they would select for leisure reading.

In the delayed recall phase, which took place one week after the initial session, I e-mailed short-answer and multiple-choice quizzes on both topics to all subjects in order to measure delayed recall. The short-answer quiz was the same as that presented during the initial session, but the multiple-choice quiz was different. Fourteen subjects responded to the e-mail.

Findings

Learning

Because humor is a subjective phenomenon and to be sure the variable "humor" was operating, I asked subjects to rate how humorous each of the versions was on a five-point Likert scale (1=strongly agree that the chapter was humorous; 5=strongly disagree). Subjects reported the humorous versions to be significantly more humorous than the nonhumorous versions ($t(17)=4.61$, $p<.001$, $\gamma=1.09$), indicating that subjects did, indeed, perceive a difference in humor in the two pieces. This validated my subjective judgment and allowed me to proceed with analyses comparing reactions to and learning from humorous and nonhumorous versions of text.

There was no significant difference on immediate recall between subjects who had read a humorous version and those who had read a nonhumorous version ($t(18)=0.40$,

$p=.70$, $\gamma=.13$). However, differences in delayed recall between humorous and nonhumorous versions approached significance ($t(14)=1.87$, $p=.07$, $\gamma=.71$). These findings are consistent with Kaplan and Pascoe's (1977) findings that humor integrated into a lecture did not appear to significantly affect immediate recall of information presented in that lecture, but that students exposed to humor performed significantly better on items testing recall of examples presented humorously. One question these findings raise is the effect of a longer delay period than I had used; specifically, what is the optimal delay for measuring the effects of humor on memory of the content of a single presentation of humorous instructional material? In addition, I made no attempt in my analyses to link questions on the recall quizzes to humor in the text. For the dissertation, I explored the link between recall of specific items and use of humor in the text.

I asked subjects what the two most important ideas were that they had learned from each chapter. Several of the subjects reported that getting ideas through observation was an important idea from the experimental ideas chapter, regardless of which version they read. There seemed to be a slight tendency for readers of the humorous experimental ideas chapter to recall affective items (e.g., students do not need to be afraid of trying to come up with experimental ideas), while readers of the nonhumorous version remembered more cognitive items (e.g., can extend other research, ideas do not have to be complicated to be important); however, the small sample size (and my own biases, which may have influenced interpretation) causes me to question the reliability of these results. Two of the eight subjects who read the humorous version remembered a humorous acronym (ROT) that referred to the importance of experimental questions being repeatable, observable, and testable, while none of the eight subjects who read the nonhumorous version listed this as one of the two most important ideas they had learned.

When asked to list the two most important ideas from the meaningfulness chapter, readers of the humorous version appeared to recall more ideas relating to the importance of making instruction meaningful (e.g., valuing students' input, the teacher's role in helping make things meaningful), while readers of the nonhumorous version seemed to recall more techniques (e.g., how to motivate students, how to reinforce information). Again, the small sample size raises the question of the relationship between humor and learning, but does not provide clear answers. The larger sample size for the dissertation research enabled me to explore the types of information subjects remember and their relationship to humor.

Interest

Subjects reported that they found the humorous versions to be significantly more interesting than the nonhumorous versions ($t(17)=2.56$, $p=.02$, $\gamma=.60$); however, expression of interest in reading more about the topic was not significantly higher for the humorous versions ($t(17)=1.00$, $p=.33$, $\gamma=.24$). Interest dropped significantly for both the humorous ($t(17)=3.00$, $p<.01$, $\gamma=.70$) and nonhumorous ($t(17)=4.37$, $p<.01$, $\gamma=1.03$) versions from before reading to after. Although it is unclear why interest dropped, this may have been due to the satisfaction of an initial sense of curiosity (Lowenstein, 1994). This effect merits further examination, but doing so is beyond the scope of this study.

The subjects' reports of interest correlated positively with test scores, although correlations with immediate recall scores were not significant. Subjects' reports of interestingness of the text and interest in the topic were correlated with delayed recall scores ($r=0.47$, $p=.12$, $r^2=.22$, and $r=0.45$, $p=.15$, $r^2=.20$, respectively). This is consistent with Alexander, Kulikowich, and Jetton's (1994) assertion that interest can stimulate depth of processing and so enhance subject-matter learning. Although it is unclear why this was more evident for the delayed recall than for the immediate recall, it may be that the effects of recency are sufficient to override any additional effects of depth of processing. Research on why this occurs goes beyond the scope of this study.

Enjoyment

The pilot study did not include a direct measure of subject enjoyment. My dissertation research asked directly about enjoyment of each reading, and I asked subjects which of the two readings they enjoyed more.

Other Effects of Humor

Subjects found the humorous versions to be significantly more informative ($t(17)=2.38$, $p=.03$, $\gamma=.56$) than the nonhumorous versions, and subjects felt they had learned more from the humorous texts ($t(17)=2.50$, $p=.023$, $\gamma=.59$). I have no explanation for these effects. I attempted to replicate these findings with my dissertation work; however, determining how or why humor affects subjects' perceptions of how informative text is and how much they learned from it was beyond the scope of this study.

There was also a tendency (though nonsignificant) to view the humorous text as more personally relevant ($t(17)=1.88$, $p=.08$, $\gamma=.44$). There have been reports in the literature that conversational text and use of the word "you" can increase perceived relevance (Means et al., 1997). The fact that humor seemed to increase perceived relevance when compared to text that used these other techniques may be worth exploring at some point in the future; however, it was beyond the scope of this study.

Subjects felt that humor supported the information it was trying to convey (mean=1.9; 1=strongly agree, 5=strongly disagree). Subjects disagreed that humor detracted from the information it was trying to convey (mean=4.1), that they would have found the materials more instructional if they had not been funny (mean=4.1), or that they would have learned more if the text had not attempted to be humorous (mean=4.2). Combined with the findings that subjects recalled more in the delayed recall quizzes, these findings led me to believe that humor may affect student learning. As based on previous research and shown in Figure 2-3 on page 10, humor may affect learning through any of a variety of mechanisms, including increasing attention, relevance, interest, and enjoyment. I designed this dissertation research to test the relationships among these variables.

Subjects, in general, reported that they found it easier to "stick with" the chapter on getting experimental ideas, although this was influenced somewhat by humor. The reason for this difference is not clear, although I suspect it may have something to do with the amount and type of information being conveyed: the chapter on getting experimental ideas covered a fairly broad range of topics, compared to the chapter on making instruction meaningful, which covered a smaller number of topics in considerably more depth.

Possible Confounding Factors

There appeared to be some effects of the topics presented on subject recall and recognition of the material. For the pilot, I prepared two versions of each of two different chapters. For one reading, I made minor modifications to a humorous chapter from an experimental psychology text (i.e., getting experimental ideas) for the humorous version, and performed a "humorectomy" (i.e., removed the humor) for the nonhumorous version. For the other reading, I used a nonhumorous chapter on making instruction meaningful that was written in a conversational style and attempted to "humorize" it. This process involved reading through the text carefully and recording humorous comments as they occurred to me. I then met with one or more others and we went through the readings paragraph by paragraph, brainstorming possible humorous additions to each paragraph. Making the text humorous was a struggle, and my perception was that some of this "humorization" may have come across as forced. (People who read the text prior to the pilot study confirmed this perception; however, the text was not changed significantly at that time, since these readers confirmed my subjective belief that the humorous version was more humorous than the nonhumorous version.)

In addition, the text for the reading on meaningfulness went into considerably more depth on a rather narrow issue, while the experimental ideas reading covered a fairly broad area in less depth. The meaningfulness chapter may therefore have provided a heavier cognitive burden on the subjects and made it more difficult for them to remember without more concentrated study of the material. Average test scores for subjects reflect this difficulty: subjects scored an average of 56 percent on the test for experimental ideas immediately after reading, and 30 percent on the meaningfulness chapter. Subjects scored 39 percent on the test of experimental ideas after one week, vs. 23 percent for the chapter on meaningfulness.

The reading on experimental ideas also contained humorous drawings that were not reproduced in equivalent form in the nonhumorous text. I do not have any feel for how these illustrations may have affected student learning. For the dissertation research, I will remove all illustrations so all effects of humor come from the text and not from illustrations.

Summary

In summary, there appeared to be several relatively positive benefits of humorous text compared to text written in an informal, conversation style: greater recall in the delayed condition, greater perception of personal relevance, and increased student interest. Although not all of these findings were statistically significant, these tendencies in such a small sample size made me believe this was a study worth pursuing. I have listed the major and minor research questions and related research questions for the dissertation study on the following pages.

Major Research Questions

The major research questions reflect what I believe to be significant effects of instruction: learning, interest in the topic, and enjoyment of the instruction. Figure 3-1 shows a model of the hypothesized relationships among the main variables.

1. Do learners recall more facts and recognize more concepts from a humorous version than from a nonhumorous version of text immediately after reading?

    1a. How does humor's proximity to tested items relate to subject recall and recognition of that information?

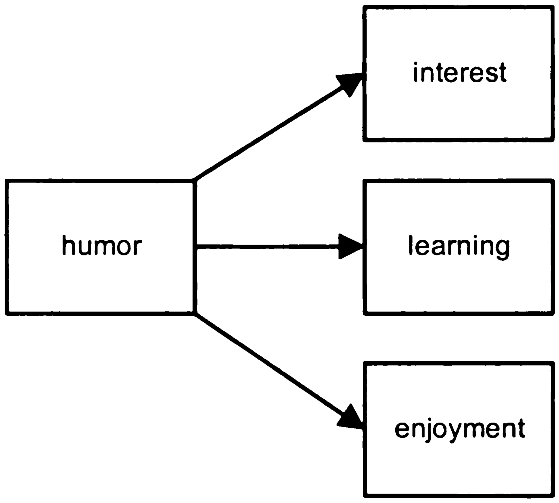

Figure 3-1. Hypothesized effects of humor in instructional

2. Do learners recall more facts and recognize more concepts from a humorous version than from a nonhumorous version of text two weeks after reading?

    2a. How does humor's proximity to tested items relate to subject recall and recognition of that information?

3. Do learners report greater enjoyment of the humorous version than the nonhumorous version?

4. Do learners express greater post-reading interest in the topic presented in the humorous version than in the topic presented nonhumorously?

Minor Research Questions

The minor research questions explore the mechanisms by which humor affects learning. Figure 3-2 shows the model with the secondary variables included.

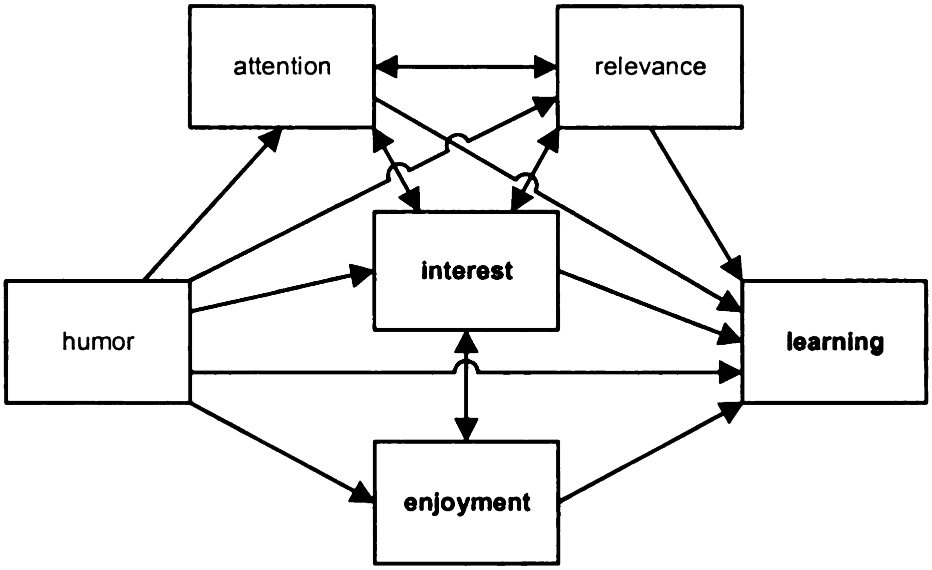

Figure 3-2. Hypothesized effects of humor in instructional text on both primary (i.e., learning, interest, and enjoyment) and secondary (i.e., attention and relevance) variables

1. Which version will subjects report held their attention better?
2. Do subjects perceive humorous material to be more personally relevant than nonhumorous material?
3. Is interest related to enjoyment?
4. Is interest related to self-reported attention?
5. Is perceived relevance related to self-reported attention?
6. Is relevance related to interest?
7. Is self-reported interest correlated with subjects' scores on the quizzes?
8. Is enjoyment related to learning?
9. Is self-reported attention related to learning?
10. Is perceived relevance related to learning?

Related Research Questions

The related research questions are not illustrated in the model (i.e., Figure 3-2) but are related to learner reactions to the use of humor in instructional text, and to other potential effects of humor.

1. Do subjects perceive humorous material to be more informative than nonhumorous material?
2. How do subjects feel humor affects the information presented (e.g., does humor support or detract from the information)?
3. Do subjects believe they learn more from humorous materials than from nonhumorous materials?
4. Do subjects spend more or less time reading humorous materials?

Model

Figure 3-3 shows the hypothesized relationships among the variables. Solid lines indicate relationships that are supported by the literature. Dotted lines indicate possible relationships (either suggested but not directly supported by the literature, and/or suggested by the pilot study). Major and minor research questions are indicated by the letters M and m, respectively, along with the number of the question. For example, the dotted line connecting humor to learning is a hypothesized relationship, and tests major questions 1 and 2.

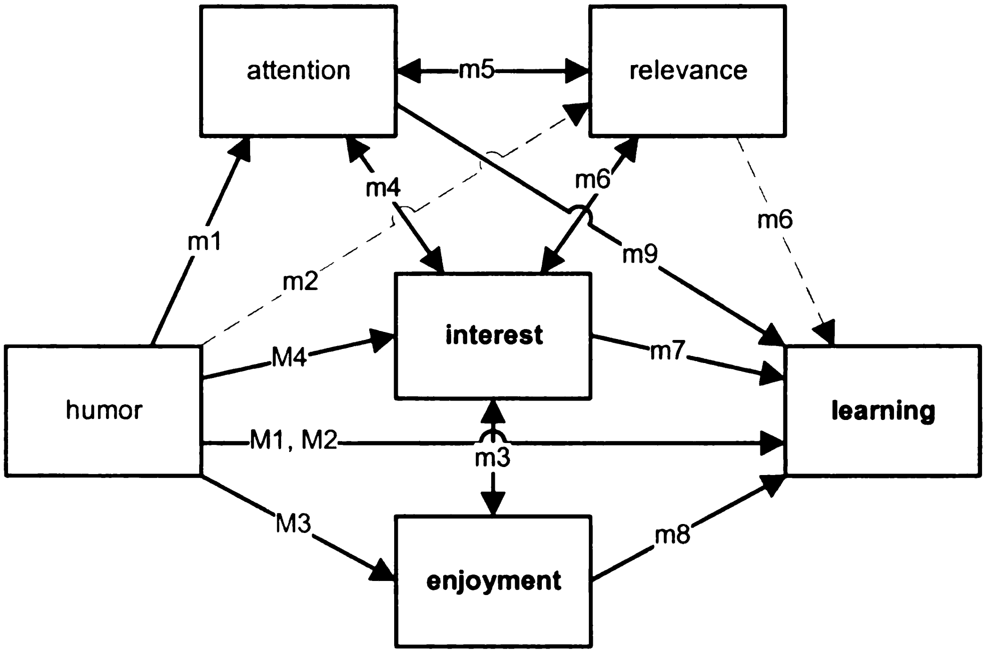

Figure 3-3. Hypothesized relationships among variables, with major and minor research questior indicated

Of my three outcome variables (i.e., learning, interest, and enjoyment), only learning appears as an endpoint in the model. Although I am interested in humor's effects on learner interest and enjoyment, I also hypothesize that these variables are mediating variables that will affect learning. These three variables appear in boxes with a thicker outline.

The figure is not a comprehensive model of all the variables that may influence humor's effect on learning. For example, humor may affect the amount of effort readers are willing to expend, and it may enable learners to better tap into their prior knowledge of a topic by increasing positive affect (Alexander, Kulikowich, and Jetton, 1994). However, I felt that adding additional variables to the model would substantially increase the complexity of the study, without contributing enough to the field's understanding of humor's effect on my three main variables to justify their inclusion.

Dissertation Study

Subjects

I was able to gain access to the following courses in order to recruit subjects for my research:

- TE150, Reflections on Learning, an introductory course in the department of teacher education.
- CEP260, Dynamics of Personal Adjustment
- CEP 261, Substance Abuse
- PSY255, Industrial/Organizational Psychology

I arranged to recruit subjects during either the first or last five minutes of class, as arranged with the instructors in advance. During this time, I gave a brief description of the study and handed out additional information about the study and a list of available times. I also informed students that the study was not limited to education majors or to students in education courses (or, in the case of the psychology course, to majors or students in psychology), and invited them to sign up with friends or roommates. By recruiting in this manner, I hoped to attract a more diverse subject pool than I would have gotten had only education students enrolled in these courses participated.

A power analysis indicated that I would need a subject pool of approximately 80 subjects, assuming $\alpha = .05$, $\beta = .7$, and an effect size of one-quarter (i.e., .25) of a standard deviation. This effect size is considered small according to Cohen (1992).

All subjects were required to be native English speakers, since the informal writing style or use of humor could be confounded by cultural variables in which it may be considered inappropriate to address the learner directly or in which the use of multiple meanings of words may be lost, as might occur in a play on words. Subjects who participated in the pilot study were not allowed to participate in the dissertation study.

A total of 133 subjects started participation in the study, and 126 completed the study. One subject was dropped from the study because she discussed answers to the delayed quiz with another subject (this second subject was dropped for other reasons). Of subjects who completed the study, 21 subjects were dropped from the analyses because they did not find the chapter that was intended to be humorous as more humorous. This criteria for dropping subjects was decided before data collection began. I considered alternatives, such as analyzing subjects according to the chapter they found more humorous rather than the chapter that was intended to be humorous; however, while humor in text can be manipulated, subjects' perceptions of that humor cannot be. I also considered analyzing subjects according to which chapter was intended to be humorous, regardless of which chapter they found more humorous; however, I decided to conduct this study under conditions as close to ideal as possible, specifically, with subjects who perceived the humor as it was intended to be perceived. Once the effects of humor on this "ideal" group are determined, I can begin to look more at individual differences in subjects' perceptions of humor and the effects of those differences on the outcome variables.

This left me with usable data from 104 subjects. Of these 104 subjects, 103 were between the ages of 18 and 24, 100 had been students continuously since graduating from high school, 103 were undergraduates, and 41 were education majors. Fifty-three of the subjects read the humorous version of the experimental ideas chapter and the nonhumorous version of the scientific fairness chapter. The remaining 51 subjects read the humorous version of the scientific fairness chapter and the nonhumorous experimental ideas chapter.

I conducted this study in accordance with policies of the University Committee for Research Involving Human Subjects (UCRIHS). Subjects provided informed consent prior to participating in the study and were debriefed after completion of the study.

Materials

My original plan was to present all chapters (i.e., readings) via paper and all quizzes and questionnaires via computer. However, Murphy's Law was in full force: the software I selected to present the quizzes was unable to support such a large project (i.e., 176 questions with 16 branching options). As I was attempting to get the computer program to work, I pursued an alternative means of presenting this information via computer (i.e., hiring a web programmer to develop a program to present the questions and collect the data), which also failed to work out. Given these roadblocks, I resorted to the computer-age equivalent of stone knives and bearskins (also known as "paper and pencil") to collect the data.

Instructional Modules

The instructional modules provided a means for manipulating the independent variable, humor. In order to separate the effect of humor from the effect of an informal, conversational tone of voice, I developed two sets of instructional materials—a humorous version and a nonhumorous version for each of two topics—in order to provide the counterbalancing described in the section on research design (see page 50). Readings were drawn from David W. Martin's *Doing Psychology Experiments*, specifically, the chapters titled "How to Get an Experimental Idea" and "How to Be Fair With Science." I received permission from Dr. Martin and from the publisher to use these sections of the book.

The experimental ideas chapter was already humorous, and few modifications were required to prepare the humorous version for the pilot study. (See Appendix A for humorous [version F] and nonhumorous [version C] versions of this reading.) For the nonhumorous version of this chapter, the humor was removed and, where possible, replaced with nonhumorous information. The author's informal, conversational tone was retained in the nonhumorous version.

The scientific fairness chapter was written in an informal, fairly conversational tone. The text itself did not contain much humor, so with the help of others, I added humor to the chapter (see Appendix B for humorous and nonhumorous versions).

I attempted to keep the humorous and nonhumorous versions of each topic the same length by having humorous material correspond with nonhumorous material; however, I was not always be able to find appropriate replacement material for the nonhumorous versions. Because of this, the humorous versions were approximately 10 percent longer than the nonhumorous versions. The humorous version of the experimental ideas chapter was approximately seven pages of 10-point type presented in two columns. The humorous chapter on being fair with science was slightly over eight pages.

Quizzes

The quizzes were developed to answer the question: Do learners learn more from humorous versions of text? I created one set of short-answer questions and two sets of multiple-choice questions for each topic. I developed questions that:

- represented what I would want students to know about the ideas presented in the chapter,
- asked about information that was likely to be remembered by readers,
- were not too difficult (i.e., they did not require a memory for trivial details),
- were not too easy (i.e., they could not be answered by simply reading the headings), and
- addressed the main ideas.

I used the same short-answer questions for both immediate and delayed recall, since these questions focused predominantly on facts. Also, since short answer questions are uncued, there was little likelihood that subjects would recall the answers they provided in the immediate recall condition when they completed them again for the delayed recall condition.

Because it seemed that there was greater likelihood of subjects remembering their answers to the multiple-choice questions, my original plan was to use two sets of multiple-choice questions for each chapter, and randomly assign subjects to one of the versions for the immediate recognition condition and to the other version for the delayed condition. I developed several fact questions, which were randomly assigned to the two versions. I also developed matched questions about concepts covered in the chapters. These questions were then randomly assigned to the two versions, with one question of each pair assigned to each version.

I ran reliability analyses on these instruments after collecting data from 25 subjects. Despite random assignment of the questions to the two versions of the quizzes, for both topics one version of the quiz was highly reliable (coefficient α = .72 and .63 for version B of experimental ideas and version A of scientific fairness, respectively) and the second version was a dud (coefficient α = -.41 and .33 for version A of experimental ideas and version B of scientific fairness, respectively). Items with a corrected item total correlation less than .30 were dropped from the quizzes, and remaining items were aggregated into a single quiz for each topic. (See Appendix C for quiz questions and answers for both chapters.) Although it was possible that subjects would remember the questions and their answers from the immediate tests of recall when they were given the delayed test, my committee and I decided that this was far preferable to continuing to use instruments that lacked reliability.

Questionnaires

The questionnaires were designed to provide answers to the questions regarding whether students find humorous text more interesting and enjoyable than nonhumorous text (research questions M3 and M4). Even if I were to find no effect of humor on learning, I might have found that humorous text was more interesting and enjoyable for students.

I developed questionnaires for the pilot test to gather information on the following:

1. subject demographics, including year in college, age, major, and number of years since high school graduation that the subject was not enrolled in courses, in order to determine the diversity of my subject pool. I also asked subjects if they were born and raised in the United States, since

cultural factors may influence perception of humor, particularly its appropriateness in instructional text.

2. initial reactions to the topics regarding how relevant and interesting subjects found each topic. These questions were intended to help me determine humor's effect on perceived relevance and on subject interest in the topic.

3. input on the presentation of the materials, including a rating of the humorousness of each chapter.

In the pilot study, the questionnaires asking for subjects' initial reactions to the topics and to the presentation of the material included only a single item for each construct (attention, relevance, etc.). In order to be able to determine reliability for these constructs, I had to create at least three items for each construct. I used the items from the pilot study as a base (e.g., "I am interested in this topic"), and then wrote additional items using synonyms or similar wording (e.g., "I expect this chapter to be interesting") or by using antonyms and reversing the scale (e.g., "I expect this topic to be boring"). Reliability analyses for each scale are reported in chapter 4.

Research Design

Subjects read the humorous version of one topic and the nonhumorous version of the other topic. Two combinations and four permutations of materials were possible:

- humorous experimental ideas/nonhumorous fairness,
- humorous fairness/nonhumorous experimental ideas,
- nonhumorous experimental ideas/humorous fairness, and
- nonhumorous fairness/humorous experimental ideas.

The order of presentation of these chapters to subjects was randomized.

The design was a within-subjects design, with each subject participating in both experimental and control conditions. This design allowed subjects to compare the humorous and nonhumorous materials on such aspects as which reading they found more enjoyable and which they would select for leisure reading. Although this design introduced the possibility of an interaction (i.e., content x humor), this problem did not arise in the pilot study. (An interaction would be a problem because it would leave me unable to interpret main effects, such as an overall effect for humor.) Assuming no interactions, this design would allow me to collapse the data across subjects and compare the effects of humorous materials to those of nonhumorous materials in the statistical analysis.

This design raised the possibility of contamination or cuing—after subjects read one chapter and before they read the next chapter, they were asked questions about how humorous they found the chapter. These questions may have cued subjects that humor was one of the variables being studied. However, with the counterbalanced design, I expected this cuing to affect all groups equally. I felt that the enhanced power provided by a within-subjects design outweighed the potential negative effects of cuing.

Procedures

I developed the materials so subjects could read all text and answer all questions within about two hours. Two hours was chosen as a limit because 1) I did not think it was reasonable to expect subjects to be able to concentrate for a period longer than that, 2) two hours allowed me to present a reasonable amount of material, and 3) two hours seemed like a fair amount of time for subjects to spend for the compensation I was able to offer.

I recorded the time it took each subject to read each chapter. This enabled me to determine reading speed for humorous and nonhumorous materials and determine if students spent more or less time reading humorous materials. Since Lorch, Lorch, and Klusewitz (1993) found that college students reported reading academic material more slowly than leisure material, I expected a comparison of the time it took subjects to read the humorous and nonhumorous chapters to provide an indication of whether the humorous text tapped into aspects of leisure reading.

I conducted the study in conference rooms and classrooms on campus. Subjects participated in the study in groups of up to 12, but subjects worked independently. Although conducting the study with different numbers of subjects in each group introduced an additional element of variability, I was forced to work within the constraints of the system. Despite my attempts to schedule subjects in groups of eight, some time periods were much more convenient to subjects and, where possible, I accommodated additional students at these times. Since subjects were randomly assigned to the four conditions, I expected that any effects of the number of participants at a single time would affect all groups equally.

Materials and questionnaires were presented via paper to subjects sequentially in the following order:

1) consent form,
2) questionnaire on demographics and initial reactions to topics,
3) one chapter,
4) questionnaire on presentation,
5) questionnaire on use of humor if chapter presented was humorous (no questionnaire specific to the nonhumorous version was presented),
6) short-answer quiz,

7) multiple-choice quiz,

8) another chapter,

9) questionnaire on presentation,

10) questionnaire on use of humor if chapter presented was humorous,

11) short-answer quiz,

12) multiple-choice quiz, and

13) questionnaire on a comparison of the two topics.

Approximately 10 days after the first session, I sent subjects an e-mail telling them that I would be sending the session 2 questions via e-mail and giving them the option of returning to the study site to complete session 2 in person. Two weeks after the first session, I sent subjects the session 2 questions via e-mail. The purpose of this second quiz was to explore delayed recall and long-term retention of ideas in the materials. Although Kaplan and Pascoe (1977) found significant differences between humorous and nonhumorous conditions six weeks after presentation, I chose a delay period of two weeks for two reasons: 1) in my pilot study, differences between humorous and nonhumorous groups approached significance using a delay period of one week, and 2) I anticipated that a six-week delay period would prove unwieldy—such a delay would likely put the delayed recall period into or after exam week for some students, and students might even forget that they had agreed to take part in a study. Unfortunately, a savvy test-taker could determine the answers to some of the short-answer questions from the multiple-choice questions. For this reason, the short-answer items were sent first and in a separate e-mail than the multiple-choice items, and I requested that subjects complete the short-answer items before opening the e-mail with the multiple-choice questions.

Subjects received $25 for participating in this study, with $10 paid in cash after the first session and $15 paid after completion of the delayed recall quiz. Subjects who participated in session 2 in person were paid in cash; subjects who completed session 2 via e-mail were paid by check. This amount of compensation averages out to about $10 an hour, which was comparable with subject compensation in other research projects.

CHAPTER 4: DATA ANALYSIS

Scale Reliability

This research used 11 subscales to determine subject attributes on the following dimensions: pre-reading interest, pre-reading perceived relevance, perceived informativeness of chapters, attention, personal interest, situational interest, perceived use of humor in the chapters, post-reading perceived relevance, enjoyment, self-perceptions of learning, and perceptions of the humor as supportive of text. Each subscale contains three or four items, as shown in Table 4-1. Subjects rated each item on a five-point Likert scale, with 5 indicating strong agreement and 1 indicating strong disagreement. Items were scaled so a high number (i.e., 5) indicated more of the attribute being measured and a low number (i.e., 1) indicated less of the attribute. Items that were reverse-scaled (so that a high number indicates more of the attribute) are indicated with "(R)."

Scale items for enjoyment and situational interest were modified from the Intrinsic Motivation Inventory for text material (Deci & Ryan, 2000). Other scales were developed by writing items similar to those that were used in the pilot test. Although these items had face validity, I did not determine construct validity for the items. I used the reliability function on the Statistical Package for the Social Sciences (SPSS) computer program, version 10.0, to determine the alpha reliability of each of these subscales. The alpha reliability provides an estimate of the amount of variance observed that is attributable to variance in the subjects' true scores (Crocker & Algina, 1986). I tested the reliability of each of these scales after collecting data from the first 25 subjects to determine if changes were needed to any of the subscale items, and again after collecting data from all subjects. With the exception of perceived learning, no modifications were needed to any of these subscales based on the first 25 subjects. Dropping one question ("I already knew the information in this chapter") from the perceived learning subscale resulted in an acceptable reliability. Table 4-1 shows the questions and the alpha reliabilities for each subscale. In order to derive a single score for each subject on each subscale, I averaged the ratings for each subject for the items on each scale. Alpha reliabilities were calculated separately for the two chapters and appear in italics for the experimental ideas chapter and in bold for the scientific fairness chapter.

Table 4-1. The eleven subscales used in this research, with the items representing those subscales indicated below each subscale name and the alpha reliabilities for the experimental ideas and scientific fairness chapters both for the first 25 subjects and for all subjects

	Alpha Reliability (*experimental ideas*, **scientific fairness**)	
Subscales	Initial (after 25 subjects)	Final
Pre-reading interest • I am interested in this topic. • I expect this topic to be boring. (R) • I expect this chapter to be interesting.	*.8017* **.6473**	*.8710* **.8478**
Pre-reading perceived relevance • This topic is relevant to me. • I expect the information in this chapter to have some value for me. • I expect the information in this chapter to be beneficial for me.	*.8188* **.8295**	*.8110* **.7845**
Perceived informativeness • This chapter was informative. • This chapter had a lot of information. • This chapter conveyed very little information relative to the amount of text. (R)	*.7739* **.6925**	*.7573* **.5141**

Table 4-1 (cont'd).

Attention • I found it difficult to pay attention to the material in this chapter as I read it. (R) • This text held my attention. • It was easy for me to concentrate on this chapter.	*.9350* **.9593**	*.8714* **.8956**
Individual (personal) interest • I would like to read more about this topic. • Now I am interested in this topic after reading this text. • I would like to know more about this topic.	*.7356* **.9262**	*.9125* **.9245**
Situational interest • This chapter was boring. (R) • This chapter was interesting. • I would describe this material as interesting.	*.8239* **.8786**	*.9015* **.9028**
Perceived use of humor in the chapters • This chapter was humorous. • I think the writer of this chapter has a good sense of humor. • I thought this chapter was funny.	*.9417* **.9626**	*.9259* **.9365**
Post-reading perceived relevance • This topic was relevant to me. • The information in this chapter will be beneficial to me. • The information in this chapter has some value to me.	*.7889* **.8537**	*.8471* **.8711**
Enjoyment • I enjoyed reading this chapter. • I liked reading this chapter. • This chapter was fun to read.	*.9085* **.9339**	*.9225* **.9292**

Table 4-1 (cont'd).

Self-perceptions of learning • I already knew the information in this chapter. (R) • This chapter was written in a way that made it easy for me to learn. • I found it difficult to learn from this chapter. (R) • I learned a substantial amount from this text.	*.7031* **.7928**[1]	*.6523* **.7791**
Humor supportive of text (humorous chapter only) • The humor in this text supports the information it's trying to convey. • The humor in this piece detracts from the information it's trying to convey. (R) • I would have found this piece more instructional if it had not used humor. (R) • I would have learned more if this text had not attempted to be humorous. (R)	*.7492* **.9034**	*.8364* **.8424**

I determined actual learning (as compared to perceived learning) through a composite score of recall (i.e., open-ended) and recognition (i.e., multiple-choice) items for each subject for each reading for both immediate and delayed recall. In other words, there were four learning measures for each subject:

- immediate measures of learning of ideas in the experimental ideas chapter
- delayed measures of learning of ideas in the experimental ideas chapter
- immediate measures of learning of ideas in the scientific fairness chapter
- delayed measures of learning of ideas in the scientific fairness chapter

[1] Initial reliability for this scale was .5255 and .6809 for experimental ideas and scientific fairness, respectively. With one item deleted from the scale, reliability was .7031 and .7928 for experimental ideas and scientific fairness, respectively.

A Note on the Analyses

I had designed this study so that subjects would serve as their own controls by reading the humorous version of one topic and the nonhumorous version of the other topic. My intention was to collapse the data across the topics so I could look at the effects of humorous text vs. nonhumorous text on the dependent variables (e.g., personal and situational interest, enjoyment, attention, perceived relevance). However, as I started to examine the data, it appeared that there were differences in subjects' perceptions of the two topics. I decided to examine the quantitative data first as two separate studies to determine if there were differences between the two topics, and if not, collapse across the topics. I found that the subjects' reactions to the variables listed above differed sharply between the two topics, which meant I could not collapse the topics. Rather than analyzing this as a 100-subject, within-subjects experiment, I analyzed it as two 50-subject between-subjects experiments.

Although I had hypothesized that humor would result in increases in the dependent variables, I felt it would be irresponsible of me as a researcher to fail to consider the possibility that humor may have detrimental effects on these variables. For this reason, I conducted all analyses as two-tailed tests rather than one-tailed tests in order to be able to determine significant adverse effects of humor, in addition to significant positive effects.

Analysis of the data in this study involved running multiple analyses. This large number of statistical analyses increased the possibility of type I errors (since approximately one in 20 analyses will result in a statistically significant result due to chance). In order to reduce the likelihood of type I errors, I used a Bonferonni correction, which involved dividing the alpha level by the number of null hypotheses that could be considered to be in the same "family," or measuring similar ideas. For example, for major research question 1a (i.e., How does humor's proximity to tested items relate to subject recall and recognition of that information?), there are five independent variables for the learning dependent variable, so each null hypothesis was tested at $\alpha/5=.01$.

In addition to providing the value of the statistics below (e.g., the t value for t-tests and the r value for correlations), I have provided the significance level (p),and the effect size. For the t-tests, the effect size is noted as γ and was calculated as the mean divided by the standard deviation. The γ value corresponds with the following effect sizes: small, $\gamma=.20$; medium, $\gamma=.50$; and large, $\gamma=.80$ (Welkowitz, Ewen, & Cohen, 2000). For the correlations, the effect size is noted as r^2, and was calculated by squaring the r value. R^2 corresponds to the following effect sizes: small, $r^2=.01$; medium, $r^2=.09$; and large, $r^2=.25$ (Cohen, 1992).

Descriptive Statistics of Scales

Tables 4-2 and 4-3 show the mean, standard deviation, possible range, and actual range for each of the scales used in these analyses, for the knowledge quizzes, and for the reading time per page.

Table 4-2. Means, standard deviations, and ranges for experimental ideas scales, including both humorous (H) and nonhumorous (NH) versions

Experimental Ideas Scales	Version	Mean	Standard Deviation	Possible Range	Actual Range
Pre-reading interest	H	3.44	.67	1-5	2-5
	NH	3.63	.70		1-5
Pre-reading relevance	H	3.45	.69	1-5	2-5
	NH	3.68	.66		2-5

Table 4-2 (cont'd).

Reading time per page	H NH	3.18 2.90	.56 .55		1.48-4 1.77-2.58
Informative	H NH	3.94 3.90	.54 .50	1-5	1.67-5 2.33-5
Attention	H NH	3.94 3.64	.58 .77	1-5	2.33-5 2-5
Personal interest	H NH	3.25 3.23	.81 .86	1-5	2-5 1.33-5
Situational Interest	H NH	3.90 3.59	.57 .81	1-5	2-5 1.33-5
Relevance	H NH	3.42 3.58	.64 .74	1-5	2-5 2-5
Perceived Learning	H NH	3.97 3.82	.44 .59	1-5	2.67-5 2.33-5
Enjoyment	H NH	3.74 3.37	.68 .85	1-5	2-5 1-5
Total Score—Immediate	H NH	16.22 14.18	4.49 3.48	0-28	2.5-23.5 7.5-21
Total Score—Delayed	H NH	11.53 10.80	3.84 3.10	0-28	3-19.5 3.5-14.5
Humor Supportive of Text	H	3.08	.59	1-5	2-5

Table 4-3. Means, standard deviations, and ranges for scientific fairness scales, including both humorous (H) and nonhumorous (NH) versions

Scientific Fairness Scales	version	Mean	Standard Deviation	Possible Range	Actual Range
Pre-reading Interest	H NH	3.44 3.39	.75 .67	1-5	1-5 2-4.33
Pre-reading Relevance	H NH	3.64 3.62	.74 .52	1-5	2-5 2.33-4.67
Reading time per page in minutes	H NH	3.53 3.41	.65 .69		2.55-5.7 2.3-5.54

Table 4-3 (cont'd).

Informative	H	3.84	.54	1-5	2.33-4.67
	NH	3.94	.54		1.67-5
Attention	H	3.12	.88	1-5	1.33-5
	NH	3.94	.58		2.33-5
Personal Interest	H	2.72	.85	1-5	1.33-5
	NH	3.25	.81		2-5
Situational Interest	H	3.26	.90	1-5	1-4.67
	NH	3.90	.57		2-5
Relevance	H	3.21	.82	1-5	2-5
	NH	3.42	.64		2-5
Perceived Learning	H	3.55	.81	1-5	1.33-5
	NH	3.97	.44		2.67-5
Enjoyment	H	3.14	.91	1-5	1.33-5
	NH	3.74	.68		2-5
Total Score—Immediate	H	19.86	6.65	0-40	7-33.5
	NH	19.87	7.88		3.5-36.5
Total Score—Delayed	H	17.68	6.19	0-40	7-30.5
	NH	18.04	6.08		6-29.5
Humor Supportive of Text	H	3.76	.73	1-5	1.5-5

Major Research Questions

My research question was: Does humor in instructional text affect learning, interest, and enjoyment? My major research questions address humor's effect on each of these three variables (i.e., learning, interest, and enjoyment). Table 4-4 summarizes the results of the major research questions. Briefly, humor had differential effects on learning, interest, and enjoyment depending on the topic.

Table 4-4. Summary of the statistical analyses of major research questions for each of the chapters

Research Question	Experimental Ideas	Scientific Fairness
M1. Does humor affect learning (immediate)?	Yes—increase	No
M2. Does humor affect learning (delayed)?	No	No
M3. Does humor affect enjoyment?	Yes—increase	Yes—increase
M4. Does humor affect interest?	Yes—increase (situational int.) No (personal int.)	Yes—increase (situational int.) No (personal int.)

Effect of Humor on Learning (Immediate)

My first major question was: Do learners recall more facts and recognize more concepts from the humorous version than from the nonhumorous version immediately after reading? Learning was measured by open-ended and multiple-choice tests. As stated in chapter 3, only those students who perceived the humor as intended (i.e., who responded that they found the humorous version to be more humorous than the nonhumorous version) were considered in the analysis. Learners recalled significantly more information from the humorous version of the experimental ideas chapter than from the nonhumorous version, and the effect size was moderate ($t(90)=2.442$, $p=.017$, $\gamma=.51$). However, there was no significant difference in recall of the different versions of the scientific fairness chapter ($t(102)=.004$, $p=.997$, $\gamma=.00$).

A related question was: How does humor's proximity to tested items relate to subject recall and recognition of that information? For this analysis, specific instances of humor had to be identified and related to the questions in the quizzes. Each of the readings was rated by two people for the humor; like the subjects, each rater read the humorous version of one chapter and the nonhumorous version of the other chapter.

Raters were two women and two men; one man and one woman had bachelor's degrees, and one man and one woman had bachelor's degrees in progress. Raters were each given the humorous version of one topic and the nonhumorous version of the other topic, and asked to underline passages that they perceived had at least attempted humor. Raters wrote a number next to each passage they underlined: a three if they found the passage very humorous, a two if they found it humorous, and a 1 if they recognized that an attempt had been made at humor, but unsuccessfully. Since what people find funny varies from person to person, I wasn't surprised that ratings of humor differed between the two raters of each chapter. In order to relate the quiz questions to the chapters, I examined those questions (or specific answers to open-ended questions) that could be linked to passages that were 1) rated as humorous by both of the raters who read the humorous chapter and 2) not rated as either humorous or attempts at humor by either of the raters who read the nonhumorous versions. I also examined questions/answers for passages that were rated humorous by all raters (regardless of humorous or nonhumorous version) or that were rated nonhumorous by all raters in order to determine if answers to specific questions were affected by overall humor in the text, rather than just specific instances of humor related to ideas in the text. Table 4-5 shows the questions and/or answers that fulfilled these criteria and were therefore compared for readers of the humorous and nonhumorous versions:

Table 4-5. Questions and answers related to text ideas that had clear ratings of use of humor by raters

Question	Answers	Average Rating of Related Humorous Text	Average Rating of Nonhumorous Text
Martin listed nine reasons why students have trouble coming up with experimental ideas. What are they?	Belief that you have to be a genius	2	0
	Belief that ideas must be completely original	0	0
	Fear of using apparatus	2.5	0
	Belief that ideas must sound scientific	3	3[†]
What is vicarious observation?	Getting ideas by reading other researchers' work	0	0

There was no significant difference in recall of either "geniuses" or "apparatus" (i.e., those answers that were related to humorous material in the humorous version only) between readers of the humorous and nonhumorous versions ($t(102)=-.94$, $p=.35$, $\gamma=-.20$ and $t(102)=1.16$, $p=.25$, $\gamma=.22$, respectively). There was no significant difference in recall of either "original ideas" or the definition of vicarious observation (i.e., those answers that were related to nonhumorous material in both version of the text) between readers of the humorous and nonhumorous versions ($t(102)=.55$, $p=.58$, $\gamma=.10$ and $t(102)=1.81$, $p=.07$, $\gamma=.36$, respectively). However, there was a significant difference between the two groups on recall of "sounding scientific," which was related to humorous text in both versions of the chapter ($t(102)=3.32$, $p=.001$, $\gamma=.69$); readers of the humorous version were significantly more likely to list "sounding scientific" as one of their answers.

[†] Two sentences were removed from this passage of the humorous version in the development of the nonhumorous version. These two sentences were rated as very humorous by both raters of the humorous version. However, the remainder of this passage, which was not rated as either humorous or as an attempt at humor by the raters of the humorous version, was rated as very humorous by both of the raters of the nonhumorous passage.

There was not clear agreement among the raters regarding use of humor in the scientific fairness chapter: in many cases, only one rater of the humorous version would perceived humor, while the other would perceive either an attempt at humor or no humor at all. Alternatively, raters of the humorous version would agree on the humor, but one or both of the raters of the nonhumorous version would detect humor in the equivalent passage in the nonhumorous version. For this reason, no analyses of individual questions/answers were performed for the scientific fairness chapter. Despite the disagreement between raters on the particulars of humor in this chapter, raters of the humorous version identified an average of 11 instances of humor, while raters of the nonhumorous version identified an average of three instances. Accordingly, the majority of subjects rated the humorous version of the scientific fairness chapter as more humorous than the nonhumorous experimental ideas chapter.

Effect of Humor on Learning (Delayed)

The second major question had to do with delayed indications of learning: Do learners recall more facts and recognize more concepts from the humorous version than from the nonhumorous version two weeks after reading? As with immediate learning, delayed learning was measured by open-ended and multiple-choice tests. There was no significant effect of humor on delayed learning for either of the two topics (experimental ideas: $t(102)=1.055$, $p=.294$, $\gamma=.21$; scientific fairness: $t(102)=.300$, $p=.765$, $\gamma=.06$).

Again, a related question regarding humor and delayed recall was: How does humor's proximity to tested items relate to subject recall and recognition of that information? I performed the same comparisons for the delayed recall as for the immediate recall, with similar results. There was no significant difference in recall of either "geniuses" or "apparatus" (i.e., those answers that were related to humorous material in the humorous version only) between readers of the humorous and nonhumorous versions ($t(102)=-1.56$, $p=.12$, $\gamma=-.30$ and $t(102)=1.76$, $p=.08$, $\gamma=.34$, respectively). There was no significant difference in recall of either "original ideas" or the definition of vicarious observation (i.e., those answers that were related to nonhumorous material in both version of the text) between readers of the humorous and nonhumorous versions ($t(102)=.38$, $p=.71$, $\gamma=.08$ and $t(102)=-.76$, $p=.45$, $\gamma=-.15$, respectively). Although there was a significant difference between the two groups on recall of "sounding scientific" for immediate recall, with the Bonferonni correction, there was not a significant difference between readers of the humorous and nonhumorous version on delayed recall of this item ($t(102)=2.39$, $p=.02$, $\gamma=.47$).

Effect of Humor on Enjoyment

The third major question was: Do learners report greater enjoyment of the humorous version than the nonhumorous version? Enjoyment was measured as a composite (i.e., arithmetic average) of three Likert scale items (see Table 4-1). Subjects reported significantly greater enjoyment of the humorous version of both the experimental ideas chapter ($t(96)=2.440$, $p=.017$, $\gamma=.48$) and the scientific fairness chapter ($t(93)=2.839$, $p=.006$, $\gamma=.56$). Effect sizes were moderate.

Effect of Humor on Interest

The final major question was: Do learners express greater post-reading interest in the topic presented humorously than in the topic presented nonhumorously? Like enjoyment, both personal and situational interest were determined by averaging the ratings of subjects on three Likert-scale items. The humorous chapter was associated with significantly greater situational interest for both chapters, with moderate effect sizes (experimental ideas: $t(88)=2.217$, $p=.029$, $\gamma=.44$; scientific fairness: $t(102)=2.036$, $p=.044$, $\gamma=.40$). In contrast, humor had no effect on personal interest for either chapter (experimental ideas: $t(101)=.149$, $p=.882$, $\gamma=.02$; scientific fairness: $t(102)=1.313$, $p=.191$, $\gamma=.26$).

Minor Research Questions

The minor research questions explore the mechanisms by which humor affects learning through the mediating variables, and the intercorrelations among these variables. As with the major questions, I found differences in subjects' responses depending on the topic.

Effect of Humor on Other Mediating Variables

I explored two additional variables, attention and perceived relevance, as potential mediating variables between humor and learning. Table 4-6 summarizes the results.

Table 4-6. Summary of the statistical analyses of major research questions for each of the chapters

Research Question	Experimental Ideas	Scientific Fairness
m1. Does humor affect attention?	Yes—increase	No
m2. Does humor affect perceived relevance?	No	No

Effect of Humor on Perceived Attention

Attention was determined by averaging subjects' responses to three Likert-scale items (see Table 4-1). Subjects reported that the humorous version of the experimental ideas chapter held their attention better than the nonhumorous version ($t(93)=2.238$, $p=.028$, $\gamma=.44$). When the p value was rounded to two digits, the difference in reported attention between the humorous and nonhumorous versions of the scientific fairness chapter was also statistically significant, with a low to moderate effect size (t(102)=1.965, p=.052, γ=.38).

Effect of Humor on Perceived Personal Relevance

Relevance was determined by averaging subjects' responses to three Likert-scale items (see Table 4-1). Subjects did not find the humorous version to be more relevant than the nonhumorous version of either the experimental ideas chapter (t(102)=1.184, p=.239, γ=.23)or the scientific fairness chapter (t(102)=1.031, p=.305, γ=.20).

Relationships Among Mediating Variables

I made several hypotheses and provided theoretical support regarding certain relationships among the mediating variables—that is, relevance, attention, situational and personal interest, and enjoyment. Table 4-7 shows my findings regarding the hypothesized relationships.

Table 4-7. Summary of the statistical analyses of major research questions for each of the chapters

Research Question	Experimental Ideas	Scientific Fairness
m3. Is interest related to enjoyment?	Yes—positive	Yes—positive
m4. Is interest related to self-reported attention?	Yes—positive	Yes—positive
m5. Is perceived relevance related to self-reported attention?	Yes—positive	Yes—positive
m6. Is relevance related to interest?	Yes—positive	Yes—positive

Although I did not make specific hypotheses about some of the relationships among these variables, for the sake of completeness I have included correlations of each of these variables with all of the others. A score for each of these variables was determined by averaging subjects' responses to three Likert-scale items. Figures 4-1 and 4-2 show the correlations among the mediating variables for the experimental ideas and

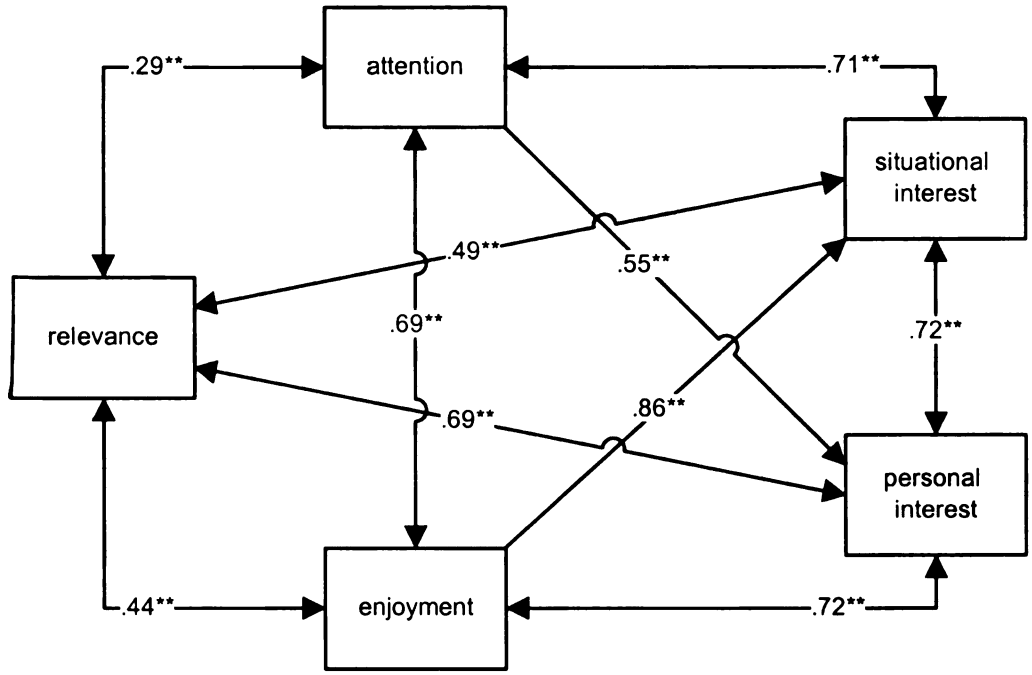

Figure 4-1. Correlations among the mediating variables for the experimental ideas chapter (**p<.01)

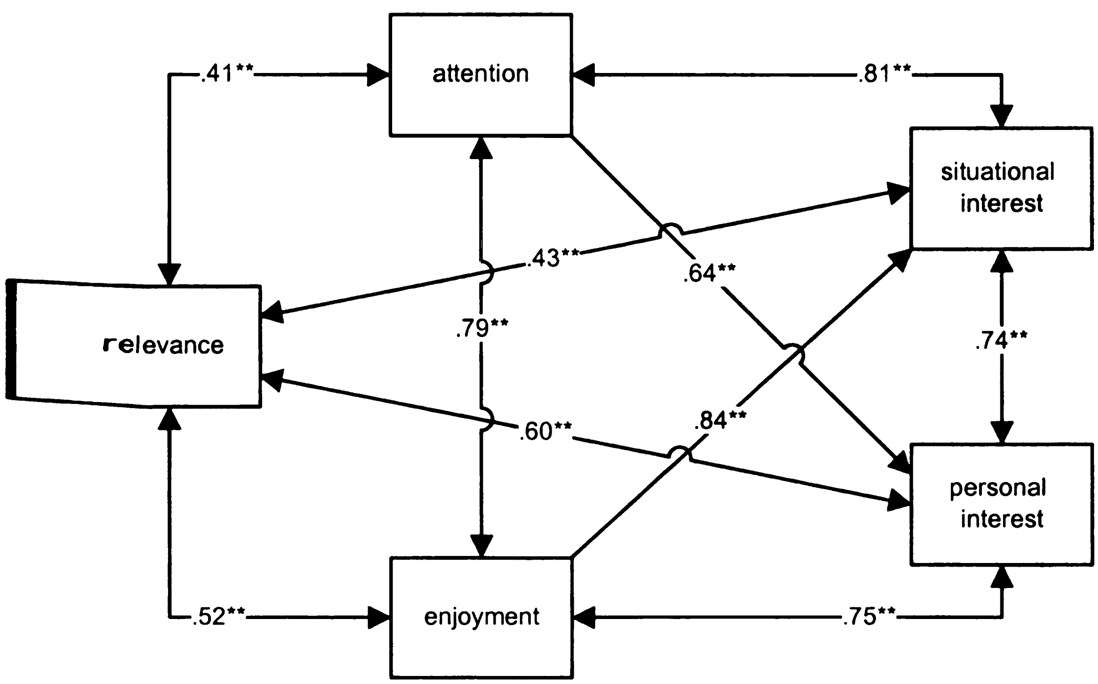

Figure 4-2. Correlations among the mediating variables for the scientific fairness chapter (**p<.01)

scientific fairness chapters, respectively. As is noted in these figures, all of the correlations were statistically significant. With the exception of the relationship between relevance and attention for the experimental ideas chapter, which was significant at $p=.003$, all of these correlations were significant at $p=.001$.

Relationship Between Perceived Relevance and Attention

Perceived post-reading relevance was significantly related to attention for both the experimental ideas chapter ($r=.291$, $p=.003$, $r^2=.085$) and the scientific fairness chapter ($r=.412$, $p<.001$, $r^2=.170$).

Relationship Between Interest and Attention

I explored this question for both types of interest (i.e., situational and personal). Both types of interest were significantly related to self-reported attention for both experimental ideas (personal interest: $r=.547$, $p<.001$, $r^2=.299$; situational interest: $r=.712$, $p<.001$, $r^2=.507$) and scientific fairness (personal interest: $r=.637$, $p<.001$, $r^2=.405$; situational interest: $r=.806$, $p<.001$, $r^2=.650$).

Relationship Between Relevance and Interest

Perceived post-reading relevance was significantly related to both situational and personal interest for both the experimental ideas (situational interest: $r=.490$, $p<.001$, $r^2=.24$; personal interest: $r=.668$, $p<.001$, $r^2=.45$) and scientific fairness (situational interest: $r=.431$, $p<.001$, $r^2=.19$; personal interest: $r=.604$, $p<.001$, $r^2=.36$) chapters.

Relationship Between Relevance and Enjoyment

Perceived post-reading relevance was positively correlated with enjoyment for both the experimental ideas chapter ($r=.442$, $p<.001$, $r^2=.20$) and the scientific fairness chapter ($r=.520$, $p<.001$, $r^2=.27$).

Relationship Between Interest and Enjoyment

Both situational and personal interest were positively correlated with enjoyment for both the experimental ideas chapter (situational interest: $r=.855$, $p<.001$, $r^2=.73$; personal interest: $r=.719$, $p<.001$, $r^2=.52$) and the scientific fairness chapter (situational interest: $r=.842$, $p<.001$, $r^2=.71$; personal interest: $r=.753$, $p<.001$, $r^2=.57$).

Relationship Between Attention and Enjoyment

Attention was positively correlated with enjoyment for both chapters (experimental ideas: $r=.690$, $p<.001$, $r^2=.48$; scientific fairness: $r=.794$, $p<.001$, $r^2=.63$).

Relationships Between the Mediating Variables and Learning

I made several hypotheses and provided theoretical support for the effects of the mediating variables on learning. Figures 4-3 and 4-4 show the correlations between each of the mediating variables on both immediate and delayed measures of learning for the experimental ideas and scientific fairness chapters, respectively. Table 4-8 summarizes the results.

Table 4-8. Summary of the statistical analyses of major research questions for each of the chapters

Research Question	Experimental Ideas	Scientific Fairness
m7. Is self-reported interest correlated with subjects' scores on the quizzes?	No	No
m8. Is enjoyment related to learning?	No	No
m9. Is self-reported attention related to learning?	No	No
m10. Is perceived relevance related to learning?	No	No

Relationship Between Interest and Learning

Neither situational interest nor personal interest was correlated with immediate learning for the experimental ideas chapter, and personal interest was not correlated with

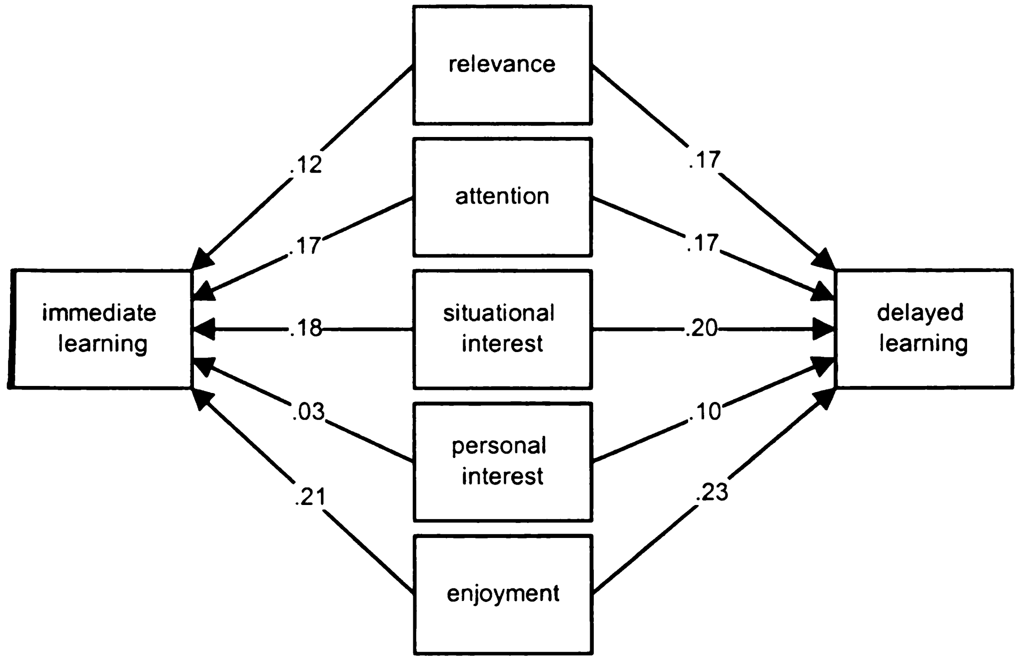

Figure 4-3. Correlations between the mediating variables and measures of immediate and delayed learning for the experimental ideas chapter

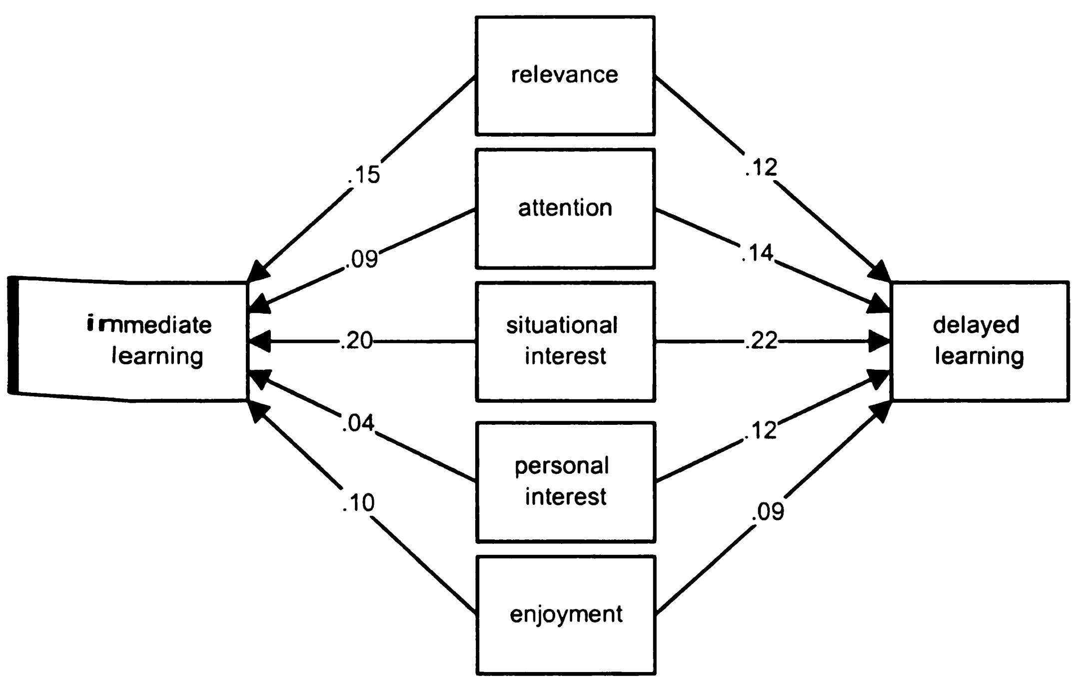

Figure 4-4. Correlations between the mediating variables and measures of immediate and delayed learning for the scientific fairness chapter

delayed learning. Although the correlation between situational interest and delayed learning was statistically significant ($r=.202$, $p=.043$, $r^2=.041$), the effect size is too low for this correlation to be considered to have any practical significance. Table 4-9 shows the values of these statistical analyses.

Table 4-9. Correlations between learning (both immediate and delayed) and interest (both personal and situational) for the experimental ideas chapter

Experimental Ideas	Immediate	Delayed
Personal interest	$r=.029$, $p=.785$, $r^2=.001$	$r=.098$, $p=.323$, $r^2=.010$
Situational interest	$r=.175$, $p=.098$, $r^2=.030$	$r=.202$, $p=.043$, $r^2=.041$

For the scientific fairness chapter, personal interest was not associated with either immediate or delayed tests of learning. Although situational interest was positively associated with both immediate and delayed tests of learning, as with the experimental ideas chapter, the effect sizes are too low for these correlations to be considered of practical significance. Table 4-10 shows the values of these statistical analyses.

Table 4-10. Correlations between learning (both immediate and delayed) and interest (both personal and situational) for the scientific fairness chapter

Scientific Fairness	Immediate	Delayed
Personal interest	$r=.042$, $p=.675$, $r^2=.002$	$r=.115$, $p=.245$, $r^2=.013$
Situational interest	$r=.200$, $p=.042$, $r^2=.040$	$r=.221$, $p=.024$, $r^2=.049$

Relationship Between Enjoyment and Learning

Enjoyment was significantly related to both immediate and delayed learning for the experimental ideas chapter; however, effect sizes were too low for these correlations to be considered of any practical significance. There was no significant relationship between enjoyment and learning for the scientific fairness chapter. Table 4-11 shows the values of these statistical analyses.

Table 4-11. Correlations between learning (both immediate and delayed) and enjoyment for the experimental ideas and scientific fairness chapters

Chapter	Immediate	Delayed
Experimental Ideas	$r=.210$, $p=.043$, $r^2=.044$	$r=.231$, $p=.019$, $r^2=.053$
Scientific Fairness	$r=.106$, $p=.348$, $r^2=.011$	$r=.093$, $p=.348$, $r^2=.009$

Relationship Between Attention and Learning

Self-reported attention was not correlated to immediate or delayed tests of learning for either of the chapters. Table 4-12 shows the values of these statistical analyses.

Table 4-12. Correlations between learning (both immediate and delayed) and attention for the experimental ideas and scientific fairness chapters

Chapter	Immediate	Delayed
Experimental Ideas	$r=.171$, $p=.099$, $r^2=.029$	$r=.172$, $p=.080$, $r^2=.030$
Scientific Fairness	$r=.088$, $p=.374$, $r^2=.008$	$r=.140$, $p=.156$, $r^2=.020$

Relationship Between Perceived Relevance and Learning

Perceived relevance was not related to immediate or delayed tests of recall for either of the chapters. Table 4-13 shows the values of these statistical analyses.

Table 4-13. Correlations between learning (both immediate and delayed) and relevance for the experimental ideas and scientific fairness chapters

Chapter	Immediate	Delayed
Experimental Ideas	$r=.115$, $\rho=.269$, $r^2=.013$	$r=.171$, $\rho=.082$, $r^2=.029$
Scientific Fairness	$r=.152$, $\rho=.123$, $r^2=.023$	$r=.118$, $\rho=.233$, $r^2=.014$

Related Research Questions

The related research questions are not illustrated in the model (i.e., Figure 3-1) but are related to learner reactions to the use of humor in instructional text, and to other potential effects of humor.

The first related research question asked if subjects perceived humorous material to be more informative than nonhumorous material. Subjects did not perceive that humorous materials were more informative than nonhumorous materials (experimental ideas: t(102)=.406, p=.685, γ=.07; scientific fairness: t(100)=.066, p=.947, γ=.013).

The second question asked how subjects felt humor affected the information presented (e.g., does humor support or detract from the information?). Overall, subjects reported that humor supported the information presented, although they found the humor more supportive in the experimental ideas chapter (mean=4.08; 1=strongly disagree, 5=strongly agree) than in the scientific fairness chapter (mean=3.76).

The third question asked if subjects believed they learned more from humorous materials than from nonhumorous materials. Subjects believed they learned more from the humorous version of the scientific fairness chapter than from the nonhumorous version (t(102)=2.389, p=.019, γ=.47). Subjects did not believe they learned more from the humorous version of the experimental ideas chapter than from the nonhumorous version (t(93)=1.475, p=.144, γ=.29). Subjects perceptions of learning, however, did not correlate significantly with actual learning for either the experimental ideas (r=.143, p=.169, r^2=.020) or the scientific fairness (r=.157, p=.111, r^2=.025) chapter.

Other Findings

I explored other effects of humor in the text by monitoring reading time for each chapter subjects read, and by asking subjects to compare the two readings on a variety of aspects. I also asked subjects questions regarding what they liked and disliked about each chapter, what helped them learn, and what the two most important ideas were that they remembered from each chapter. These findings are discussed in turn below.

Reading Time

I recorded reading time for the two versions of the chapters. When adjusted for length (since humorous versions were longer than nonhumorous versions), there was no difference in the amount of time it took subjects to read humorous vs. nonhumorous text (experimental ideas: $t(97)=.163$, $p=.871$, $\gamma=.03$; scientific fairness: $t(98)=.880$, $p=.381$, $\gamma=.18$).

Comparisons Between Chapters

I asked four questions that had subjects compare the two readings. I asked which chapter subjects found it easier to stick with, which chapter they enjoyed more, which chapter they believed was more likely to be in a textbook, and which chapter they would choose for leisure reading. Tables 4-16, 4-19, 4-22, and 4-25 show the frequencies of subjects' responses to these questions. In these tables, the two chapters read by subjects are presented in a column (e.g., subjects who read the humorous version of the experimental ideas chapter, as shown at the top of the column, selected either that chapter or the nonhumorous version of the scientific fairness chapter).

More Likely in Textbook

When asked which chapter they felt they would be more likely to find in a textbook, most subjects said they would be more likely to find the nonhumorous version in a textbook: 88 percent of subjects who read the humorous experimental ideas (HEI) chapter and nonhumorous scientific fairness (NSF) chapter reported that the latter would be more likely to appear in a textbook. Sixty-six percent of subjects who read the nonhumorous experimental ideas (NEI) chapter and the humorous scientific fairness (HSF) chapter reported that the former would be more likely to appear in a textbook.

Table 4-14 shows the frequencies of subjects who said each chapter would be more likely to be found in a textbook.

Table 4-14. Number of subjects who selected each response when asked which of the chapters they would be most likely to find in a textbook

Chapter more likely to find in textbook	Chapters Read	
	HEI/NSF	NEI/HSF
Experimental ideas	6	31
Scientific fairness	45	16
Both	2	4
Neither	0	1

The main reasons cited by subjects that a certain chapter would be found in a textbook had to do with either the content generally (e.g., "More information that related to something people might study in a course"), the use of humor (e.g., "Textbooks don't usually use humor to help instruction"), or the overall presentation (e.g., "Textbooks are generally boring and formal"). Additional comments from subjects were not clear enough to be classified (e.g., "The text was very similar.") Frequencies of responses are shown in Tables 4-15 and 4-16. Subjects were allowed to list more than one reason, so total responses may add up to more than 104.

Table 4-15. Frequencies of subjects' reasons why a chapter would be more likely to appear in a textbook (readers of the humorous version of the experimental ideas chapter and nonhumorous version of the scientific fairness chapter)

Reasons subjects listed believing the chapter would appear in a textbook	HEI	NSF
Presentation—general	0	25
Presentation—humor	0	3
Content	2	12

Table 4-16. Frequencies of subjects' reasons why a chapter would be more likely to appear in a textbook (readers of the humorous version of the scientific fairness chapter and nonhumorous version of the experimental ideas chapter)

Reasons subjects listed believing the chapter would appear in a textbook	HSF	NEI
Presentation—general	10	21
Presentation—humor	0	4
Content	7	4

As shown in the tables above, subjects cited humor as a reason a chapter would *not* be more likely to be found in a textbook.

Easier to Stick With

When asked which chapter was easier to stick with, most subjects (65 percent) reported that the humorous chapter was easier to stick with. However, preferences varied between topics: 92 percent of the subjects who read the humorous experimental ideas chapter found it easier to stick with that chapter than the nonhumorous scientific fairness chapter; in contrast, 37 percent found it easier to stick with the humorous scientific fairness chapter than the nonhumorous experimental ideas chapter. Table 4-17 summarizes the information; subjects who read the same two chapters appear in the columns.

Table 4-17. Number of subjects who selected each response when asked which of the chapters was easier to stick with

Chapter selected as easier to stick with	Chapters Read	
	HEI/NSF	NEI/HSF
Experimental ideas	49	32
Scientific fairness	4	19

Subjects' responses for most of the open-ended opinion questions could be divided into the following categories:

- humor (e.g., "It was more fun to read with the humor" [subject #101]),

- interest (e.g., "The topic was more interesting" [#167]),
- easy to read or understand (e.g., "It was written more clearly—it kind of sounded like someone talking to you in a conversation" [#161]),
- use of examples (e.g., "He used examples … to restate his ideas" [#120]),
- enjoyment (e.g., "It was more fun to read" [#208]),
- relevance or ability to relate to (e.g., "It was information that could be [beneficial] to anyone even if you are not a scientist" [#121]),
- attention (e.g., "Got my attention early on and kept it" [#106]), and
- read first (e.g., "I read this one first, so I was a little more awake" [#136]).

Tables 4-18 and 4-19 show the break-down of reasons student gave for the chapter they found easier to stick with.

Subjects who read the humorous version of the experimental ideas chapter cited a variety of reasons for finding that chapter easier to stick with than the nonhumorous scientific fairness chapter, with humor, interest, and ease of reading/understanding cited most frequently (see Table 4-18). Subjects who read the humorous scientific fairness chapter were most likely to cite the humor as the reason they found that chapter easier to stick with, but subjects who preferred the nonhumorous experimental ideas chapter cited interest and ease of reading/understanding (see Table 4-19). Four students said they found the first chapter they read to be the one they found easier to stick with; this is not surprising, given the amount of concentration required to complete the study (i.e., subjects read two chapters and answered over 100 questions in a two-hour period).

Table 4-18. Frequencies of subjects' reasons why the chapter they selected was easier to stick with (readers of the humorous version of the experimental ideas chapter and nonhumorous version of the scientific fairness chapter)

Reasons subjects listed for finding their preferred chapter easier to stick with	HEI	NSF
Humor	24	1
Interesting	15	0
Easy to Read/Understand	14	0
Examples	2	0
Enjoyable	3	0
Relevant/Related To	9	0
Held Attention	3	0
Read First	1	0
Smaller segments	2	1

Table 4-19. Frequencies of subjects' reasons why the chapter they selected was easier to stick with (readers of the humorous version of the scientific fairness chapter and nonhumorous version of the experimental ideas chapter)

Reasons subjects listed for finding their preferred chapter easier to stick with	HSF	NEI
Humor	12	2
Interesting	2	8
Easy to Read/Understand	6	10
Examples	5	2
Enjoyable	0	1
Relevant/Related To	1	0
Held Attention	1	0
Read First	2	1

The tables show only those responses indicated by at least two students. Several comments were listed by only one subject each. Comments by subjects who preferred the humorous experimental ideas chapter were: it was to the point, seemed shorter, had a broader range, used less jargon, and was more organized (one student each). Other comments by students who preferred the nonhumorous experimental ideas chapter were: it was more common sense, the topic was interesting, the subject paid more attention after being quizzed after the first chapter, there was less "excess material," experimental ideas are fun to think about, there was more activity in the material, and the explanations were short. Another comment by a student who preferred the humorous scientific fairness chapter was that it was more complicated. Other comments by students who preferred the nonhumorous scientific fairness chapter were: knowing where the chapter was going, humor didn't detract from the information, and the information was new for the subject.

More Enjoyable

The majority of subjects reported that they found the experimental ideas chapter more enjoyable, regardless of which version of that chapter they read: 94 percent of subjects said they enjoyed the humorous experimental ideas chapter more than the nonhumorous scientific fairness chapter, and 67 percent enjoyed the nonhumorous experimental ideas chapter more than the humorous scientific fairness chapter. Table 4-20 shows the frequencies of subjects who identified each of the chapters as more enjoyable.

Table 4-20. Number of subjects who selected each response when asked which of the chapters they enjoyed more

Chapter selected as more enjoyable	Chapters Read	
	HEI/NSF	NEI/HSF
Experimental ideas	49	34
Scientific fairness	3	17

Subjects who read the humorous version of the experimental ideas chapter cited a variety of reasons for finding that chapter more enjoyable than the nonhumorous scientific fairness chapter, with humor, interest, and ease of reading/understanding cited most frequently. Readers who found the humorous scientific fairness chapter more enjoyable cited the humor, while readers who enjoyed the nonhumorous experimental ideas chapter more said they found it more interesting or more relevant. Tables 4-21 and 4-22 show subjects' written responses when asked why they found their chosen chapter more enjoyable.

Table 4-21. Frequencies of subjects' reasons why they found the chapter they selected more enjoyable (readers of the humorous version of the experimental ideas chapter and nonhumorous version of the scientific fairness chapter)

Reasons subjects listed for finding their preferred chapter more enjoyable	HEI	NSF
Humor	20	0
Interesting	8	0
Easy to Read/Understand	8	0
Examples	3	0
Relevant/Related To	13	0
Held Attention	4	0
Organization	0	1

Table 4-22. Frequencies of subjects' reasons why they found the chapter they selected more enjoyable (readers of the humorous version of the scientific fairness chapter and nonhumorous version of the experimental ideas chapter)

Reasons subjects listed for finding their preferred chapter more enjoyable	HSF	NEI
Humor	9	2
Interesting	2	8
Easy to Read/Understand	3	4
Examples	4	4
Relevant/Related To	1	15
Held Attention	1	1

Other comments listed by students who preferred the humorous experimental ideas chapter were: it was interesting to realize the observations that could be made every day, it involved interacting with the general public more, it flowed better, it was faster to read, and the information in the scientific fairness chapter was obvious. Other comments listed by students who preferred the nonhumorous experimental ideas chapter were: it wasn't overwhelming, it was helpful, it "gave creative ... ways to do what was being explained," there was more activity, and it didn't have as much useless information. Other comments listed by students who preferred the nonhumorous scientific fairness chapter were that the information was new (to the subject) and better organized. There were no additional comments listed by students who preferred the humorous scientific fairness chapter.

Leisure Reading

When asked which chapter they would choose for leisure reading, subjects again preferred the experimental ideas chapter over the scientific fairness chapter. Ninety-six percent of subjects who read the humorous experimental ideas chapter preferred it over the nonhumorous scientific fairness chapter, while 35 percent of the subjects who read the humorous scientific fairness chapter preferred it over the nonhumorous experimental ideas chapter. Table 4-23 shows the frequencies of subjects who identified each of the chapters as the one they would select for leisure reading.

Table 4-23. Number of subjects who selected each response when asked which of the chapters they choose for leisure reading

Chapter selected for leisure reading	Chapters Read	
	HEI/NSF	NEI/HSF
Experimental ideas	51	33
Scientific fairness	2	18

Subjects who had read the humorous version of the experimental ideas chapter cited reasons such as interest, ease of reading, relevance, and the chapter's ability to hold their attention, in addition to humor. About one-third (35 percent) of subjects who read the humorous version of the scientific fairness chapter said they would select it for leisure reading. Several of these subjects cited the humor and enjoyment of the chapter as reasons for their choice; reasons for selecting the nonhumorous experimental ideas chapter include interest and relevance. Tables 4-24 and 4-25 show the answers written by subjects when asked why they would choose their preferred chapter for leisure reading.

4-24. Frequencies of subjects' reasons why they would prefer the chapter they selected for leisure reading (readers of the humorous version of the experimental ideas chapter and nonhumorous version of the scientific fairness chapter)

Reasons subjects listed for selecting their preferred chapter for leisure reading	HEI	NSF
Humor	11	0
Interesting	10	1
Easy to Read/Understand	12	0
Enjoyable	1	0
Relevant/Related To	8	0
Held Attention	9	0
Informative	0	1

4-25. Frequencies of subjects' reasons why they would prefer the chapter they selected for leisure reading (readers of the humorous version of the scientific fairness chapter and nonhumorous version of the experimental ideas chapter)

Reasons subjects listed for selecting their preferred chapter for leisure reading	HSF	NEI
Humor	8	0
Interesting	0	11
Easy to Read/Understand	3	4
Examples	1	0
Enjoyable	6	5
Relevant/Related To	1	8
Held Attention	2	1
Informative	3	2

Other comments listed by students who preferred the humorous experimental ideas chapter were: it was common sense, the subject wouldn't be as concerned about retaining the information, the subject liked the topic more, the topic was more helpful, it was more entertaining, it was less like a textbook, it seemed more like a self-help type of reading, and the scientific fairness chapter was too obvious. Additional comments listed by students who preferred the nonhumorous experimental ideas chapter were: it was more comfortable or pleasant to read, the reading moved steadily from one idea to the next, it would make the subject observe the little things they do, it required less concentration, and the subject liked the material better. There were no additional comments listed by students who preferred either version of the scientific fairness chapter.

Responses to Each of the Chapters

Subjects were asked what they liked and disliked about each of the chapters and what helped them learn the information in the chapters. For ease of comparison, responses to the humorous and nonhumorous versions of each chapter are presented together in each table below. Tables 4-26 and 4-27 show subjects' reports of what they liked about the humorous and nonhumorous versions of the experimental ideas chapter (Table 4-26) and the scientific fairness chapter (Table 4-27).

Elements Subjects Liked

Readers of the humorous version of the experimental ideas chapter cited humor and interest as things they liked about the chapter. Subjects who read the nonhumorous chapter cited the writing style, the fact that it was easy to read or understand, the examples, relevance or their ability to relate to the material, and organization as factors they liked about the chapter. Comments could be grouped into categories similar to those used for most of the comparisons, with the addition of the category of text characteristics (e.g., "Bold face print told you what was important").

Table 4-26. Subjects' reports regarding what they liked about the humorous and nonhumorous versions of the experimental ideas chapter

	HEI	NEI
Humor	11	0
Informative	2	3
Summary	3	2
Interesting	7	3
Easy to Read/ Understand/Writing Style	9	31
Examples	4	9
Enjoyable	0	0
Relevant/Relate To	2	5
Held Attention	2	2
Organization	2	11
Text Characteristics	2	5

Table 4-27. Subjects' reports regarding what they liked about the humorous and nonhumorous versions of the experimental ideas chapter

	HSF	NSF
Humor	26	2
Informative	3	1
Summary	2	3
Text Characteristics	1	7
Interesting	2	1
Easy to Read/Understand/ Writing Style	11	8
Examples	17	21
Relevant/Related To	3	1
Organization	6	13

About half of the readers of the humorous version of the scientific fairness chapter cited the humor when asked what they liked about the chapter. This percentage was much higher than that for the humorous experimental ideas chapter. Readers of the nonhumorous chapter were more likely to list text characteristics (e.g., use of bold or italics, headings, etc.) and organization as factors they liked about the chapter. Forty subjects cited the ease of reading or the writing style as factors they liked about the experimental ideas chapter, compared to 19 readers of the scientific fairness chapter. In contrast, examples were cited by 13 readers of the experimental ideas chapter and 38 readers of the scientific fairness chapter.

Other responses regarding what subjects liked about the humorous experimental ideas chapter were that it was easy to remember, the points were stated directly, it was brief, it addressed human needs and fears, it was written to help the reader do something, explanations were simple, it didn't try to be scientific, and it gave the subject confidence to come up with an idea. Subjects listed the way the information was broken down, the layout, and that it made them think about things they hadn't thought about before as reasons they liked the nonhumorous experimental ideas chapter.

Other responses regarding what subjects liked about the humorous scientific fairness chapter were that the ideas were short, they liked the content itself, it was straightforward, and it had a lot of information in a small amount of space. Other responses regarding what subjects liked about the nonhumorous scientific fairness chapter were miscellaneous facts and that they learned new things.

Elements Subjects Disliked

Tables 4-28 and 4-29 show subjects' reports of what they did not like about the humorous and nonhumorous versions of the experimental ideas chapter (Table 4-28) and the scientific fairness chapter (Table 4-29).

Table 4-28. Subjects' reports regarding what they did not like about the humorous and nonhumorous versions of the experimental ideas chapter

	HEI	NEI
Nothing	10	14
Text Characteristics	2	2
Dry/Boring/Not Interesting/Didn't Hold Attention	5	13
Too Long/Too Much Text for Amt of Info	2	7
Too Much Information	4	2
Too Many Examples	0	0
Lack of Examples	0	1
Humor	1	0
No Humor	0	3
Not Relevant	4	3
Repetitive	2	2
Organization	4	0
Phobias/Big Words	7	0
Author Talked Down	0	3

Table 4-29. Subjects' reports regarding what they did not like about the humorous and nonhumorous versions of the scientific fairness chapter

	HSF	NSF
Nothing	6	5
Text Characteristics	0	2
Dry/Boring/Not Interesting/Didn't Hold Attention	13	19
Too Long/Too Much Text for Amt of Info	12	10
Too Much Information	2	5
Too Many Examples	2	1
Lack of Examples	0	2

Table 4-29 (cont'd).

Humor	7	0
No Humor	0	4
Not Relevant	1	3
Repetitive	1	2
Organization	0	1
Author Talked Down	0	0

Twenty-four of the readers of the experimental ideas chapter, regardless of the version (i.e., humorous or nonhumorous), said there was nothing they disliked about it, compared to 11 of the readers of the scientific fairness chapter. Readers of the nonhumorous version of the experimental ideas chapter were more likely than readers of the humorous version to say it was dry, boring, not interesting, or didn't hold their attention.

Readers of both versions of the scientific fairness chapter felt that it was dry, boring, not interesting, or didn't hold their attention, and that it was too long or had too much text for the amount of information it contained.

One reader of the humorous version of the experimental ideas chapter said they didn't like the humor, compared to seven of the readers of the humorous scientific fairness chapter. Three of the subjects who read the nonhumorous version of the experimental ideas chapter said they disliked that it did not contain humor, while four of the readers of the nonhumorous scientific fairness chapter cited the lack of humor as something they disliked about the chapter.

Elements that Helped Learning

Subjects were asked what helped them learn from each of the chapters. Tables 4-30 and 4-31 show subjects' responses to this questions for the humorous and nonhumorous versions of the experimental ideas chapter (Table 4-30) and the scientific fairness chapter (Table 4-31).

Table 4-30. Subjects' reports regarding what helped them learn the information in the experimental ideas chapter

	HEI	NEI
Humor	14	0
Overview/Summary	2	4
Interesting	2	0
Easy to Read/ Understand/Writing Style	16	12
Examples	16	22
Enjoyable	1	0
Relevant/Relate To	6	0
Held Attention	1	0
Organization	7	23
Text Characteristics	7	4

Table 4-31. Subjects' reports regarding what helped them learn the information in the scientific fairness chapter

	HSF	NSF
Humor	13	0
Overview/Summary	3	7
Interesting	1	0
Easy to Read/ Understand/Writing Style	8	4
Examples	25	35
Relevant/Relate To	1	0
Organization	7	10
Text Characteristics	3	3

Twenty-six percent of the subjects felt that humor had helped them learn the information in the humorous chapters, and this was not dependent on which chapter they read. Examples were cited for both chapters, with 60 subjects citing the examples in the scientific fairness chapter and 38, the experimental ideas chapter. Readers of the nonhumorous versions were somewhat more likely to say the examples helped them learn. Organization was also cited frequently as something that helped subjects learn, with three times as many readers of the nonhumorous version of the experimental ideas chapter citing organization as readers of the humorous version of that chapter. For both chapters, only readers of the humorous version reported that the chapter's relevance to them or their ability to relate to the material helped them learn.

Most Important Ideas

Subjects were asked to list the two most important ideas they learned from each chapter. Table 4-32 shows subjects' responses when asked what the two most important ideas were that they learned from the experimental ideas chapter. Recalled ideas could be categorized into the following areas:

- fears and phobias (general) (e.g., "I learned the things that prevent people from using creative ideas" [subject #150],
- finding ideas is easy (e.g., "Ideas for experiments are all around" [#116]),
- ideas come from observation (e.g., "Direct observation can be a good tool when trying to form questions" [#137]),
- questions can be answered through research (e.g., "Practical questions will be answered through research" [#124]),
- how to get ideas (nonspecific) (e.g., "The ways of coming up with ideas" [#122]),
- fears interfere with ideas (e.g., "There are a lot of irrelevant fears that prevent people from doing an experiment" [#207]),
- not to be afraid (e.g., "One should not be afraid of asking creating questions" [#124]),
- observation is important for science (e.g., "Observation is the key for effective research" [#202]), and

- a good experiment should be repeatable, observable, and testable (e.g., "Characteristics of an experimental questions (ROT)" [#203]).

Table 4-32. Subjects' reports regarding the two most important ideas they learned from the experimental ideas chapter

	HEI	NEI
Fears/Phobias (general)	9	11
Finding Ideas is Easy/Everyone has Ideas	17	18
Ideas from Observation	14	13
Answer Questions Through Research	2	0
How to Get Ideas (nonspecific)	6	4
Fears Interfere with Ideas	9	4
Don't be Afraid	6	4
Observation Important for Science	3	3
Repeatable, Observable, Testable	2	4

There were no clear differences in what readers said were the most important ideas from the experimental ideas chapter.

For the scientific fairness chapter, recalled ideas could be categorized into the following areas:

- cheating hurts science and society (e.g., "Bad scientific research hurts everyone" [#120]),
- general do's (e.g., "Do experiments ethically" [#159]),
- specific do's (e.g., "Reorganize to make things clear" [#175]),
- general don'ts (e.g., "It is unethical and immoral to conduct research poorly" [#120]),
- specific don'ts (e.g., "Not to fabricate results" [#173]),
- consequences of lying (e.g., "Falsifying an experiment can cause you to lose your credibility" [#138]),
- to be skeptical of research, or that scientists sometimes lie (e.g., ""Fraud happens in all professions, even one as important as science" [#116]),

- differences among the types of "tricks" (e.g., "There are right ways and wrong ways that people can 'change' their experiments" [#209]), and
- the importance of ethics (e.g., "Scientific study and research needs to have a solid foundation of ethical and moral standards when it comes to doing research and experiments" [#133]).

Table 4-33 shows subjects' responses when asked what the two most important ideas were that they learned from the scientific fairness chapter.

Table 4-33. Subjects' reports regarding the two most important ideas they learned from the scientific fairness chapter

	HSF	NSF
Cheating Hurts Science/Society	9	9
Do's (general)	7	3
Do's (specific)	8	9
Don'ts (general)	6	7
Don'ts (specific)	14	5
Consequences	4	7
Be Skeptical/Scientists Lie	14	8
Differences Between Types of Tricks	6	9
Importance of Ethics	6	6

Readers of the humorous version appeared more likely to list specific don'ts and the need to be skeptical of scientific reports than readers of the nonhumorous version.

Summary

Obviously, there are individual differences in appreciation of humor. This was demonstrated both by differences between the raters in their ratings of the humor in the chapters, and by subjects' comments: while some subjects said they liked the humor in the humorous chapters, other subjects reported that they disliked the humor.

Despite differences in humor appreciation, analysis of the quantitative data indicated that humor can help students learn under certain conditions. Humor in instructional text can also have other positive benefits (such as enhancing interest, enjoyment, and attention). Figure 4-5 shows the hypothesized relationship among the variables when I began this study. Figures 4-6 and 4-7 show the actual relationships among these variables as indicated by this research for the experimental ideas and scientific fairness chapters, respectively.

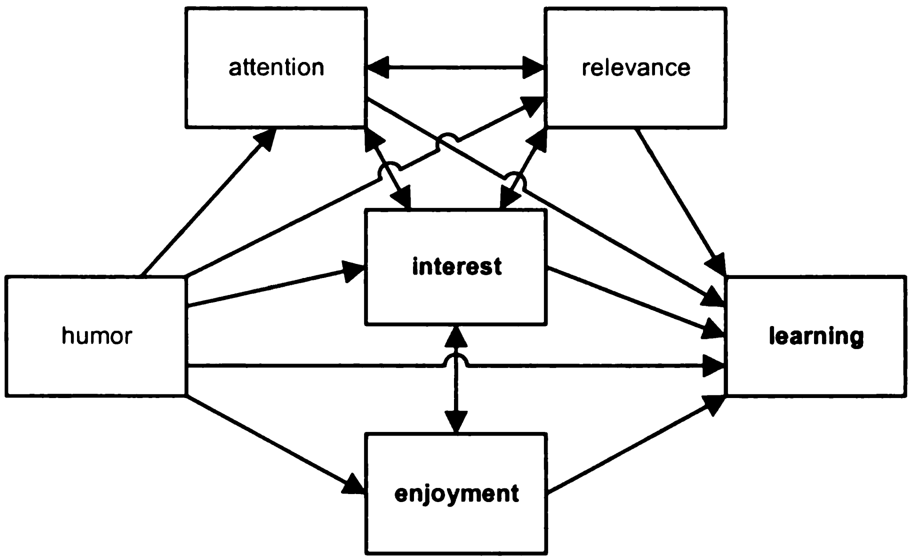

Figure 4-5. Hypothesized effects of humor in instructional text

The qualitative data suggested differences between the two chapters in the humor and the difficulty of the content. In written responses regarding what they liked about each chapter, subjects reported that the experimental ideas chapter was interesting and easy to read and understand; subjects liked the organization and examples in the scientific fairness chapter. When asked what they disliked about the readings, subjects reported that the nonhumorous version of the experimental ideas chapter and both versions of the scientific fairness chapters were dry, boring, or didn't hold their attention. Subjects also reported that the scientific fairness chapter (regardless of whether there was humor or not) was too long or had too much text for the amount of information it contained. Several readers of the humorous version of the scientific fairness chapter disliked the

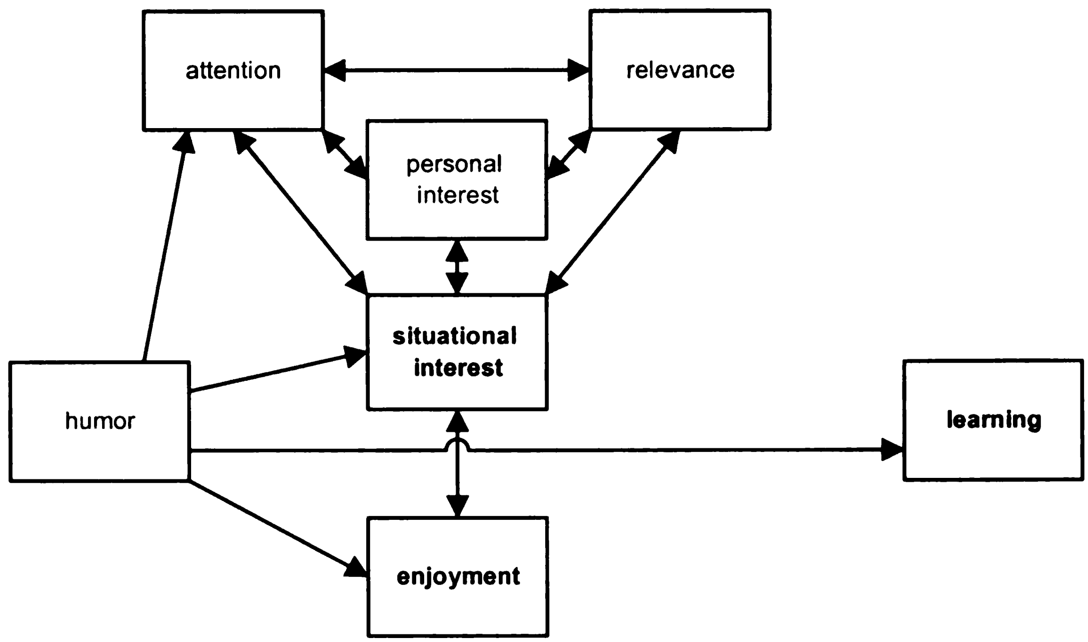

Figure 4-6. Connections among variables for the experimental ideas chapter

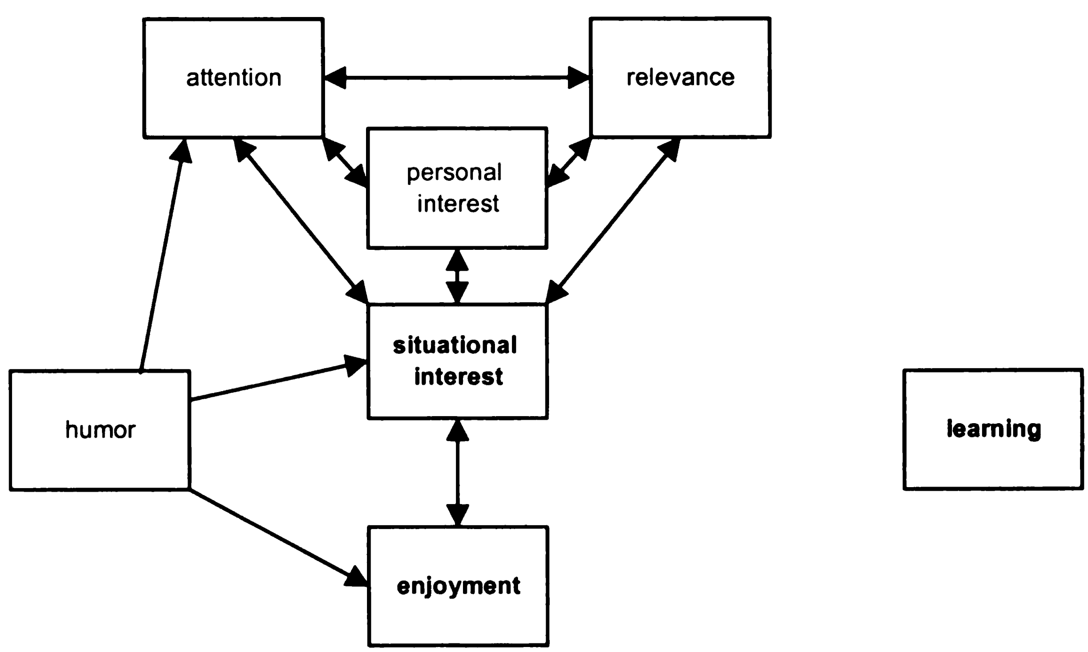

Figure 4-7. Connections among variables for the scientific fairness chapter

humor, while only one reader of the humorous experimental ideas chapter disliked the humor.

In comparisons of the chapters, subjects reported that the experimental ideas chapter (regardless of humor) was easier to stick with, more enjoyable, and would be preferred for leisure reading. Subjects reported that they would be more likely to find the nonhumorous version (regardless of topic) in a textbook.

CHAPTER 5: DISCUSSION

The purpose of this research was to determine if humor in instructional text affects learning, interest, and enjoyment. In this chapter, I review the arguments I made in the literature review regarding the effect of humor, compare my expectations to my findings, and discuss possible interpretations of my results.

Immediate Effects of Humor on Learning

My first major research question was: Do learners recall more facts and recognize more concepts from the humorous version than from the nonhumorous version immediately after reading? In chapter 2, I presented both theoretical and empirical support for the effect of humor on learning. Briefly, humor has been found to enhance learning in some classroom studies (e.g., Bryant, Comisky, Crane, & Zillmann, 1980; Ziv, 1988). Theoretical support for the effect of humor on learning comes from dual coding theory, which states that comprehension and memory of text are affected by both verbal (e.g., text, speech) and nonverbal (e.g., imagery, affect) systems (Sadoski, Goetz, & Fritz, 1993), and that the effects of these two systems are additive.

In this study, I found that immediately after reading, learners recalled significantly more information from the humorous version of the experimental ideas chapter than from the nonhumorous version, but that there was no significant difference in recall of the different versions of the scientific fairness chapter.

Why was there an effect of humor on immediate recall with the experimental ideas chapter and not the scientific fairness chapter? I have identified two possible reasons for the differences in immediate recall: 1) differences in the amount, type, and integration of humor in the two chapters, and 2) differences in the cognitive burden placed on readers of each chapter. I discuss each of these possibilities below.

Reason 1: Differences in Amount and Type of Humor

Amount of Humor

As stated in chapter 4, the humor in each version of each chapter was rated by two people. Although two raters is not ideal, it does allow me to speculate on how specific instances of humor may have been perceived by the subjects. Differences in the amount and type of humor in each chapter are reflected partially in the ratings given to the passages by these raters. There was much more agreement among the raters on the humor in the versions of the experimental ideas chapter: in the humorous version, two raters identified 14 and 16 instances of humor, respectively, with agreement between the two raters on seven passages. The two raters of the nonhumorous version of the experimental ideas chapter identified five and six instances of humor in that chapter, respectively, with agreement on two instances.

There was considerably less agreement between the raters for the scientific fairness chapter: raters of the humorous version identified 16 and six instances, respectively, with agreement on only two passages. Raters of the nonhumorous version identified six and zero instances, respectively, with no agreement on any of the humor. This lack of agreement between raters of the humor in the scientific fairness chapter may indicate that the humor in that chapter has a more individual appeal. In contrast, the relative agreement of the raters of the humorous experimental ideas chapter may indicate a more universal appeal of the humor in that chapter.

As stated in chapter 4, raters of the nonhumorous version of the experimental ideas chapter identified part of a passage as humorous that was not identified as humorous by the raters of the humorous version. However, raters of the humorous version identified sentences that surrounded that passage as humorous. (These sentences had been removed from the nonhumorous version.) The fact that the text was identified as humorous by the raters of the nonhumorous version but as not humorous by raters of the humorous version indicates that there may be a context effect for humor. In other words, text that is surrounded by nonhumorous text may be more likely to be identified as humorous than the same text that is surrounded by humorous text.

Types of Humor

In addition, the two chapters had different types of humor, due in part to the development of the chapters. The experimental ideas chapter had humor injected by the author, and few changes were made for the development of the humorous version of that chapter. The humor in this chapter drew largely from the content of the chapter itself. As an example, Martin (the author of the chapters) talked about fears students have about trying to come up with ideas for experiments. He states in that chapter, "Psychologists call irrational fears phobias. Since I am a psychologist, I cannot resist the temptation to name the phobias behind our inability to get experimental ideas. (Any resemblance of these names to accepted psychological terminology is purely coincidental.)" (Martin, 2000, p. 43). Martin proceeds to name the "phobias" using Latin and Greek (or Latin- and Greek-sounding) prefixes that are commonly associated with the fears he wishes to convey: for example, "pseudononphonoscientiaphobia" is the fear of not sounding scientific.

In contrast, the scientific fairness chapter did not have much humor originally, so humor had to be added to develop the humorous version. The humor added to this version referred more often to characters and ideas *outside* the readings, like Inspector Clouseau and Buffy the Vampire Slayer. Obviously, these references would only be understood by readers who brought with them the requisite knowledge to understand the references. This may explain why there was little agreement between the raters for this chapter: different people bring different experiences with them, and these differences may have been reflected by differential appreciation of the humor in the readings. In addition, since the humor in the scientific fairness chapter came from outside, instead of more directly from the content of the reading, the humor may have come across as more forced and possibly less related to the content. Subjects noted these differences in their comments on the humorous scientific fairness chapter: "I noticed more jokes in the article, even though they were cheesy—they made it more interesting" (subject #209), and "Neither was funny to me, but this one tried to make funny comments that I noticed" (#178).

In order to further study the differences in the types of humor used in this study, the humor must be classified into specific types of humor using a taxonomy of humor. Several taxonomies have been proposed, based on different aspects of humor. For example, Wandersee (1982) lists 15 sources of humor (such as word-play, understatement, and departure from propriety) and 16 forms (including jokes, puns, riddles, and cartoons). Long and Graesser (1988) described a taxonomy of jokes that included 10 categories (including nonsense and social satire) and a taxonomy of wit that included 11 categories (such as irony, satire, sarcasm, and understatement). Berger (1997) described four categories of humor (i.e., humor involving language, logic, identity, or action), under which 45 techniques are grouped (such as irony, absurdity, parody, and slapstick). Although in this study I have sufficed to say that different types of humor were used in the two chapters, future research should 1) categorize the types of humor used in instructional text and 2) determine the effects of these different types of humor on subjects' learning and reactions to the text.

Integration of Humor

As noted by Zillmann and Bryant (1983, cited in Robbins, 1994), humor that is well integrated with the material can lead to greater retention of the information. As stated earlier, the humor in the experimental ideas chapter drew directly from the content of the chapter, while the humor in the scientific fairness chapter referred more to "outside" situations or characters. The humor added to the scientific fairness chapter therefore may not have been sufficiently integrated with the material to produce the desired benefits.

Cohen's Theory of Humor

Cohen (1999) developed a theory that supports differences in types of humor. He describes two types of jokes: conditional jokes are those that will work only with specific audiences, and pure jokes are those which are universal because they presuppose nothing in the audience. Cohen further points out that truly pure jokes don't really exist, since they presuppose, at the very least, a familiarity with the language in which the joke is told. Conditional jokes require the audience to bring with them the knowledge or experience required from them to either get the joke or find it amusing. Cohen notes that "it is a vital feature of much joking that only a suitably qualified audience—one that can meet the condition—can receive the joke, and the audience often derives an additional satisfaction from knowing this about itself" (p. 12). The best jokes are concise, and this concision is possible because a certain amount of knowledge is assumed on the part of the audience. When I consider the two humorous chapters in light of Cohen's theory of humor, I can see that the experimental ideas chapter involved a purer form of humor, since the humor was derived largely from the text itself, while the scientific fairness chapter involved more conditional humor by requiring subjects to bring with them certain background or experiences that would enable them to understand the jokes.

Cohen's (1999) theory of humor states that humor works in part by creating a community of the listeners. Cohen proposes that "a deep satisfaction in successful joke transactions is the sense held mutually by teller and hearer that they are joined in feeling" (p. 25). A failure of a listener (or reader) to "get the joke" represents a "failure to join one another in the community of appreciation" (p. 26). By relying on conditional jokes, the joke-teller pulls the listener in by requiring the listener to supply something. This sense of community is not created if the joke is explained or if the listener doesn't "get" the joke. Community is created by letting the listener know that he or she and the joke-teller share common knowledge or experiences.

Comedians such as Rita Rudner and Eddie Izzard help to create a community by referring later in their routines to jokes made earlier—in effect, they create an entire audience that is then "in on" the (inside) joke. By referring to ideas outside the chapters, I inadvertently required that subjects bring certain knowledge and experiences with them in order to "get" the joke. Subjects who lacked the necessary knowledge or experiences to understand the humor may have felt excluded from any "community" that may exist between the author and the readers. However, since subjects did, in general, perceive the humorous chapters to be more humorous than the nonhumorous chapter, it is difficult to know exactly what the effects of this "added" humor were.

Reason 2: Differences in Cognitive Burden

In addition to the differences in the amount and type of humor in the two chapters, the content of the scientific ideas chapter may have been more difficult for students to understand. In support of this assertion, reading time was significantly longer for the scientific fairness chapter (3.46 minutes per page) than for the experimental ideas chapter (2.92 minutes per page), with a large effect size ($t=10.933$, $\rho<.001$, $\gamma=1.12$). Subjects' comments also reflected this difficulty. When asked what they disliked about the chapter, many subjects commented that it was dry, boring, or not interesting, or that it was too long or had too much text for the amount of information: "I didn't like how boring the chapter was. I found it hard to concentrate and little motivation to keep reading" (#182) and "It seemed to drag on and I felt like I was receiving too much information at one time" (#105).

Subjects may also have lacked the necessary background to be able to easily integrate this new information: "Occasionally, some of the vocabulary specific to experiments could get a little confusing and I had to read more slowly and carefully" (subject #108) and "Wasn't clear on some materials and even after I reread the section twice, it still didn't make sense" (#204). This lack of background may have required students to work harder to understand and assimilate the new information.

In addition, percentage correct on the quizzes was significantly lower for the scientific fairness chapter (50.00 percent) than for the experimental ideas chapter (54.29 percent), although the effect size was low ($t=3.023$, $\rho=.003$, $\gamma=.32$). While I can't claim that the tests for the two chapters were equivalent, taken together with the other evidence, this difference in percentage correct may support a difference in cognitive burden. There was no apparent floor or ceiling effect for either of the chapters immediately after reading, and the fact that some subjects scored quite well (e.g., high score was 84 percent for each of the chapters) indicates that it was possible for subjects to score well on the tests with only one reading.

Possibly the increased cognitive burden posed by the scientific fairness chapter made the humor inappropriate: the increased level of concentration required may mean that the humor was not able to enhance learning.

Conclusion

This combination of the differences in the type and amount of humor in these two chapters, combined with differences in the cognitive burden placed on readers of each chapter, may have contributed to the differences in effect of humor in the text on subject recall and recognition of information in the two chapters. This general finding of the effect of humor on learning is consistent with Schmidt's finding that humor facilitates learning; however, as you will see later, other of my findings contradict Schmidt's finding.

Delayed Effect of Humor on Learning

The second major question had to do with delayed indications of learning: Do learners recall more facts and recognize more concepts from the humorous version than from the nonhumorous version two weeks after reading? In my pilot study, scores on the delayed quizzes (i.e., one week after the reading) approached significance. I decided to increase the delay period for this study because Kaplan and Pascoe (1977) found significant effects of humor as measured six weeks after the event. I had hoped that the effects of humor in my study would be more pronounced with a longer delay period than the pilot study, with readers of humorous materials retaining more information than readers of the nonhumorous materials. However, humor was not associated with increased retention of material after two weeks for either of the chapters. A major difference between my study and the study by Kaplan and Pascoe is that the latter occurred in the context of a college class. As stated earlier, it may have been unreasonable to expect that a single reading of the materials (especially with a lack of motivation to remember the material or even read it carefully) would yield significant results.

The definition and measurement of learning may also have affected my results for delayed learning. Learning is defined as "getting knowledge of (a subject) or skill in (an art, trade, etc.) by study, experience, instruction, etc." or "fixing in the mind" (Agnes, 2000). Unfortunately, subjects in this study had no incentive to study or learn the material, or even to read carefully. "Learning" in this study was measured by open-ended and multiple-choice questions that were asked immediately after reading, and for practical purposes, was defined in this study as what subjects could recall from a single reading of the material. Rather than representing true learning, which results from carefully reading and studying materials, this measure of learning reflects only what students can remember after a single exposure to the material.

Effect of Humor Related to Specific Passages of Text

Immediate Recall

In addition to looking at the overall effect of humor in instructional text on learning, I also explored how recall related to specific humorous passages in the text. I found that, for information related to humor only in the humorous version of the text, there was no significant difference in recall of information between readers of the humorous and nonhumorous versions. For information related to nonhumorous material in both versions of the text, there was also no significant difference in recall of answers between readers of the humorous and nonhumorous versions. However, I did find a significant difference between the two groups on recall of information which was related to humorous text (as defined by raters) in both versions of the chapter, with readers of the humorous chapter significantly more likely to recall the information than readers of the nonhumorous chapter.

These findings were unexpected, especially when considered in light of the finding of increased overall learning for subjects who read the humorous version of the experimental ideas chapter. If recall was directly related to humor, I would expect no difference in subjects' scores on the two versions for passages that either were humorous in both versions or nonhumorous in both versions, and differences in favor of the humorous version for passages that were humorous in the humorous version and not humorous in the nonhumorous version. Given the relative lack of humor in the nonhumorous version of the chapter, this instance of humor may have been a distraction to subjects—the humor may have been so unexpected that subjects were startled or confused by it, rather than amused by it. An additional experiment involving very limited humor in one version of a reading and more humor in another version and questions related to humorous and nonhumorous passages in each, perhaps with a qualitative component to determine which passages individual subjects find humorous, may provide insight into these findings. An alternative explanation for these findings is that the raters' sense of humor is not similar to or representative of the senses of humor of my subjects.

These findings do not support Schmidt and Williams's (2001) study, which found that humor enhanced recall of specific items. The finding that specific instances of humor could not be linked to improved performance on specific test items may argue in favor of the effect of positive affect, rather than any effect of humor that comes from an increase in attention to specific instances of humor.

The humorous versions of these chapters contained considerably more humor than the amount considered ideal in a classroom (i.e., three to four instances in a 50-minute lecture). Since it took subjects 20 to 30 minutes to read each chapter, one or two instances of humor would have been consistent with the ideal amount of humor for classroom instruction; this amount is much closer to the amount of humor contained in the nonhumorous chapters. However, written instruction lacks many of the aspects of classroom instruction that may contribute to perceptions of humor, including the instructor's intonation and facial expression, the ability to tailor humor to a particular class, and an ongoing relationship between the instructor and the students. The finding that subjects who read the nonhumorous version of the experimental ideas chapter did worse on a question related to a humorous passage may indicate that the ideal amount of humor may be considerably higher in instructional text than in a classroom environment.

Delayed Recall

I also explored how humor's proximity to tested items related to subject recall of that information in delayed tests of learning. I performed the same comparisons for delayed recall as for immediate recall, and found no significant differences between groups in recall of any answers. These findings are consistent with my findings that there was no difference in overall recall after two weeks. These results do not support my hypothesis that humor would reduce the rate of forgetting.

Future Research

Future studies should look at the effect of humor in instructional materials in the context of an actual classroom setting, where students have the ability and the incentive to study, and where they have the background needed to understand the material. A qualitative study would also be helpful to determine what individual subjects find amusing and how that relates to their recall of ideas in the text. Also, more research is needed to determine if there is an optimal amount of humor in written instructional materials.

Effect of Humor on Enjoyment

Another major research question was: Do learners report greater enjoyment of the humorous version than the nonhumorous version? In chapter 2, I provided some evidence from the literature that supported a link between humor and enjoyment. However, none of that evidence was related to humor in instructional materials. This study explored the link between humor and enjoyment as it related to instructional materials. As reported in chapter 4, subjects reported significantly greater enjoyment of the humorous version of both of the chapters. This confirms my hypothesis that humor in instructional materials would increase enjoyment. As stated in chapter 2, I believe that enjoyment is an important effect of humor in instruction, and these findings support this positive effect of humor in instructional text.

Effect of Humor on Interest

The final major research question was: Do learners express greater post-reading interest in the topic presented humorously than in the topic presented nonhumorously? In the literature review, I cited empirical evidence for the connection between humor and interest in the classroom (e.g., Ziv, 1988; Bryant et al., 1980). I also cited theoretical evidence for the effect of humor on situational interest. My study supported the effect of humor on situational interest: subjects reported significantly greater situational interest in the humorous versions of the two chapters; however, there was no effect of humor on reported personal interest for either of the chapters. This finding is consistent with Schraw and Lehman's (2001) assertion that personal interest is context-general (and therefore unaffected by the style of presentation), while situational interest is context-specific and can be elicited through the use of certain devices or characteristics of the presentation.

Minor Research Questions

The minor research questions explored the mechanisms by which humor affected learning through the mediating variables, and the intercorrelations among those variables.

Effect of Humor on Attention and Relevance

I proposed that humor may aid learning by directing learners' attention and increasing perceived relevance, which would facilitate encoding and retrieval of information. Attention and relevance may also interact with interest and with each other, producing a synergistic effect and ultimately affecting learning.

Attention

Humor was associated with increased attention in both the experimental ideas and scientific fairness chapters. This finding is consistent with my review of the literature.

Relevance

In chapter 2, I noted that use of a visible author enhanced perceived relevance (Paxton, 1997) and expressed interest in finding out if humor would additionally enhance relevance, perhaps by increasing the text equivalence of instructor immediacy. I was unable, however, to provide any direct empirical or theoretical support for the effect of humor on perceived relevance, and my study did not support the effect of humor on relevance for either of the chapters. In retrospect, this finding is not surprising: there is no reason to expect that humor would increase perceived relevance.

Relationships Among Mediating Variables

In chapter 2, I noted that the writings of several researchers indicate considerable conceptual overlap among the ideas of interest (both personal and situational), attention, and perceived relevance. Correlations among all of these variables were statistically significant for both the experimental ideas and scientific fairness chapters. These findings are neither surprising nor unexpected.

Although I did not predict correlations between enjoyment and each of these variables, the high correlations, again, are neither surprising nor unexpected: we accept as common sense that we are likely to enjoy things were are interested in and believe to be relevant, and that we pay more attention to things we enjoy.

The fact that these correlations were significant for both of the readings indicates that the relationships among these constructs may be independent of topic; that is, interest, relevance, attention, and enjoyment tend to be all high or all low, depending on each individual's reaction to the topic. The correlations among these variables support a conceptual overlap among these constructs, and raises the possibility that these are not, in fact, distinct constructs. Supporting this view, a factor analysis I performed after the first 25 subjects indicated that these variables loaded on a single factor. However, supporting a contrary view, the fact that humor affects these variables differentially (i.e., increases enjoyment and situational interest, and has no effect on personal relevance) in ways that are consistent with the theories cited in chapter 2, indicates that these may be distinct constructs.

Relationships Between the Mediating Variables and Learning

I made several hypotheses and provided theoretical support for the effects of the mediating variables on learning. The fact that none of these variables correlated with learning may provide support for Schmidt and Williams's (2001) findings, which indicated that humor itself, rather than incongruity, resolution of incongruity, bizarreness, or distinctiveness, enhanced recall. Although Schmidt and Williams obtained these results in a study of memory for cartoons, my findings suggest that humor may also have an effect on recall of ideas in text.

An alternative explanation for my findings is that there is a mediating variable other than interest, enjoyment, attention, or relevance that is affected by humor and influences learning. Possibilities include the effect of physiological arousal (Vance, 1987), increased general motivation to do well, or desire to please an authority figure (such as the instructor, or possibly even the author). A qualitative study of subjects who appreciate the humor and who perform well on tests of recall and recognition may help to identify additional variables.

Relationship Between Interest and Learning

As stated in the literature review, the presence or absence of interest, whether personal or situational, influences the type of learning strategies individuals employ. Learners who are interested in what they are learning use deep-processing strategies (i.e., making connections, identifying and solving problems on their own), while learners who are not interested use surface-level strategies (i.e., memorization) (Krapp, Hidi, & Renninger, 1992). I also cited several studies that found connections between interest and learning.

My findings did not support this connection between interest and learning: neither situational interest nor personal interest was correlated with immediate learning for the experimental ideas chapter, and personal interest was not correlated with delayed learning. The correlation between situational interest and delayed learning, while statistically significant, had an effect size too low to be considered of any practical significance. Similarly, for the scientific fairness chapter, personal interest was not associated with either immediate or delayed tests of learning, and although situational interest was positively associated with both immediate and delayed tests of learning, the effect sizes were too low for these correlations to be considered of practical significance.

The way learning was defined and measured in this study may have contributed to the lack of relationship I found between interest and learning: as stated earlier, subjects read the material once and there was no opportunity to study and no incentive to read the materials carefully. Although the presence or absence of interest influences the type of learning strategies individuals employ, as stated earlier, there was no incentive for subjects to actually learn the material.

Krapp et al. (1992) argued that situational interest motivates individuals to seek out new information. Similarly, individuals who are personally interested in a topic will seek out new information on that topic. Due to limitations of the study, I don't know if humor has any long-term effects on subjects' interest, or if subjects could be more likely to seek out or at least be open to additional experiences with the topic. A longitudinal study that explores interest immediately after reading and its relationship to 1) interest after several weeks or months and/or 2) subjects' responses to additional readings on the topics after several weeks or months would provide information on the long-term effects of humor on interest.

Relationship Between Enjoyment and Learning

I proposed a dual role for enjoyment in learning: enjoyment may influence learning directly through affective mechanisms, and it may also make learners more receptive to additional experiences with the topic.

Although enjoyment was significantly related to both immediate and delayed learning for the experimental ideas chapter, effect sizes were too low for these correlations to be considered of any practical significance. There was no significant relationship between enjoyment and learning for the scientific fairness chapter.

Although a learner who enjoys the learning experience may be more receptive to, and perhaps even seek out, additional experiences with the topic, unfortunately there was no way to determine that in this study. As with the link between interest and learning, a longitudinal study that explores subjects' attitudes toward the topics several weeks later, and perhaps even their responses to additional readings on the same topic, would help determine if enjoyment can affect receptivity to additional experiences with the topic.

In chapter 2, I cited research providing support for the idea that affect influences learning. Specifically, positive affect (such as enjoyment) results in a reliance on general knowledge structures, which reduces demand on cognitive resources, which in turns frees up those resources for other tasks, such as making connections with previous learning and additional processing. The results of this study did not support the effect of positive affect on learning; my findings failed to support Isen, Daubman, and Gorgoglione's (1987) hypothesis that positive affect results in deeper processing of new information, with more connections being made to previously known information. However, as stated earlier, students had no incentive, and likely no intrinsic desire, to actually learn this information.

Relationship Between Attention and Learning

Keller (1987) argued that attention is an essential prerequisite to learning—subjects are not likely to learn anything if they are not paying attention. This study did not find that attention was related to learning: self-reported attention was not correlated with learning for either of the chapters as measured by either immediate or delayed tests of learning. Although this finding seems counterintuitive, attention is likely insufficient by itself to cause learning. In the case of this study, subjects had no opportunity or incentive to study the materials in depth, or even to read them carefully; subjects had nothing at stake (e.g., a grade or other compensation) that would be likely to increase the amount of effort they were willing to put into learning the information in the readings.

More research, preferably in a classroom setting in which students have incentive to learn, is needed to determine if there is a correlation between attention and learning.

Relationship Between Perceived Relevance and Learning

Perceived relevance may influence the amount of effort learners are willing to expend in order to learn, and subjects' perceived relevance can be enhanced in text through use of certain strategies. Despite theoretical and empirical support for the relationship between perceived relevance and learning, perceived relevance was not related to immediate or delayed tests of recall for either of the chapters. In this study, however, subjects were probably more focused on getting the task done than in doing well at the task, and they completed the task regardless of its relevance.

Related Research Questions

I explored several questions related to the role of humor in learning that do not appear in the model presented in chapter 3 (i.e., Figure 3-1).

Informativeness

The first related research question asked if subjects perceived humorous material to be more informative than nonhumorous material. Although my pilot study suggested that subjects found humorous materials to be more informative, in this larger study, subjects did not perceive that humorous materials were more informative than nonhumorous materials. Although this result is not surprising, the fact that this was not significant using a two-tailed test indicates that subjects do not perceive humorous materials to be any *less* informative than nonhumorous materials. Authors' concerns that their materials will be perceived as less informative if they use humor would appear to be unfounded, according to these results.

Humor Supportive of Text

The second question asked how subjects felt humor affected the information presented (e.g., does humor support or detract from the information?). Overall, subjects reported that humor supported the information presented, although they found the humor more supportive in the experimental ideas chapter (mean=4.08; 1=strongly disagree, 5=strongly agree) than in the scientific fairness chapter (mean=3.76). This is consistent with my earlier statements regarding the type and integration of humor into the two chapters: humor was more related to the ideas in that text and more integrated into the text itself.

Perceived Learning

The third question asked if subjects believed they learned more from humorous materials than from nonhumorous materials. Subjects believed they learned more from the humorous version of the experimental ideas chapter than from the nonhumorous version, but they did not believe they learned more from the humorous scientific fairness chapter. Oddly, subjects' perceptions of learning did not correlate significantly with actual learning for either of the chapters. A qualitative study exploring subjects' perceptions of learning as it relates to humorous versus nonhumorous materials, along with a comparison of how these perceptions relate to actual learning, might shed light on these findings.

Other Findings

Reading Time

I recorded reading time for the two versions of the chapters. When adjusted for length, there was no difference in the amount of time it took subjects to read humorous vs. nonhumorous text. This did not support my hypothesis that humorous text would tie into aspects of leisure reading as reported by Lorch, Lorch, and Klusewitz (1993). There are at least three possible explanations for this: 1) humor in instructional text does not automatically cause readers to tap into aspects of leisure reading, 2) subjects were not expected to study this text, so there was no incentive to spend any more time on the text than was absolutely necessary to "get the job done," or 3) subjects' perceptions of the amount of time it takes to read academic material compared to leisure material is inaccurate. Lorch et al. (1993) reported that students *perceived* their reading to be faster for leisure reading, even when they were reading with the intention of learning something. However, in the study by Lorch et al., there were no comparisons of actual reading speed. More research is needed to determine if leisure reading, and particularly leisure reading with the goal of learning something, is faster than reading done as part of a class assignment.

Comparisons of Two Chapters

Subjects were asked which chapter they found easier to stick with, which chapter they enjoyed more, which chapter they felt was more likely to be in a textbook, and which chapter they would choose for leisure reading. Subjects were also asked to explain their answers.

Easier to Stick With

When asked which chapter was easier to stick with, most subjects (78 percent) reported that the humorous chapter was easier to stick with. However, preferences varied between topics: 92 percent of the subjects who read the humorous experimental ideas chapter found it easier to stick with that chapter than the nonhumorous scientific fairness chapter. Similarly, 63 percent found it easier to stick with the nonhumorous experimental ideas chapter than the humorous scientific fairness chapter. Obviously, with so many subjects preferring the experimental ideas chapter, the two chapters were not equivalent in terms of ease of reading, amount and difficulty of content, and/or other factors. However, since subjects who read the humorous version of the scientific fairness chapter were more likely to select that chapter than readers who read the nonhumorous version, humor may have some ability to retain subjects' interest in informational text. In order to determine humor's ability to hold a learner's attention, more research is needed using texts that are approximately equivalent in terms of interest and difficulty of content.

More Enjoyable

Both versions of the experimental ideas chapter were enjoyed more than the scientific fairness chapters: 94 percent of subjects said they enjoyed the humorous experimental ideas chapter more than the nonhumorous scientific fairness chapter, and 67 percent enjoyed the nonhumorous experimental ideas chapter more than the humorous scientific fairness chapter. Again, since subjects who read the humorous version of the scientific fairness chapter were more likely to select it than were readers of the nonhumorous version of the scientific fairness chapters, this indicates that humor may affect readers' enjoyment, even when the material is difficult.

More Likely in Textbook

When asked which chapter they felt they would be more likely to find in a textbook, most subjects selected the nonhumorous chapter they had read: 88 percent of subjects who read the humorous experimental ideas chapter and nonhumorous scientific fairness chapter reported that the latter would be more likely to appear in a textbook.

Sixty-six percent of subjects who read the nonhumorous experimental ideas chapter and the humorous scientific fairness chapter reported that the former would be more likely to appear in a textbook.

Although 76 of the 98 subjects who responded to this question said they would be more likely to find the nonhumorous version in a textbook, only seven of the subjects said this was because of the humor. The largest percentage of subjects (47 percent) cited the presentation style more generally (e.g., "Readings out of books usually tend to be more monotonous and hard to pay attention too [sic]" [subject #123]). It is not clear why students overwhelmingly chose the nonhumorous version as the one that would appear in a textbook but did not cite humor as the reason for this, since the overall style of presentation was the same for both humorous and nonhumorous chapters (i.e., humor was removed from the nonhumorous version, but the author's casual and direct style of writing was retained). This suggests that humor may have an effect on subjects' perception of the presentation of the material that goes beyond the humor itself. A replication of this study using two chapters that are more closely matched for type and amount of humor, difficulty of content, and interest to subjects may provide a better idea of how humor affects readers' responses and impressions of instructional material.

Preferred for Leisure Reading

When asked which chapter they would choose for leisure reading, subjects again preferred the experimental ideas chapter over the scientific fairness chapter. Ninety-six percent of subjects who read the humorous experimental ideas chapter preferred it over the nonhumorous scientific fairness chapter, while 35 percent of the subjects who read the humorous scientific fairness chapter preferred it over the nonhumorous experimental ideas chapter.

As with the findings regarding the easier chapter to stick with and which chapter subjects enjoyed more, humor appears to have influenced which chapter students would select for leisure reading. Nineteen of 104 subjects cited the humor when asked why they would select their chosen chapter. Twelve subjects who read the humorous version of the experimental ideas chapter said it was easy to read and understand, compared to only four subjects who read the nonhumorous version of that chapter, and three subjects said they would select the humorous version of scientific fairness because of the ease of reading. This suggests that humor may contribute to subjects' perceptions of a reading as being easier to read.

Likes

When subjects were asked what they liked about each of the chapters, twice as many readers of the humorous version than the nonhumorous version of the experimental ideas chapter reported that it was interesting. Many more of the subjects who read the nonhumorous experimental ideas chapter cited the writing style or the fact that it was easy to read or understand, the examples, relevance or their ability to relate to the material, and organization as factors that they liked about the chapter. Only 20 percent of the readers of the humorous experimental ideas chapter cited the humor as something they liked about the chapter.

About half of the readers of the humorous version of the scientific fairness chapter cited the humor when asked what they liked about the chapter. This percentage was much higher than that for the humorous experimental ideas chapter. Readers of the nonhumorous scientific fairness chapter were more likely to list text characteristics (e.g., use of bold or italics, headings, etc.) and organization as factors they liked about the chapter. Approximately equal numbers of subjects of each version of the scientific fairness chapter listed ease of reading or the writing style as something they liked about the chapter; however, only 19 subjects cited this, compared to 40 of the readers of the experimental ideas chapter. The fact that more subjects said they liked the ease of reading or the writing style of the experimental ideas chapter than the scientific fairness chapter is consistent with findings that suggest that the scientific fairness chapter was more difficult for subjects.

Regardless of which chapter they read, more readers of the nonhumorous versions than the humorous versions reported that organization was something they liked about the chapter. More research is needed to determine if humorous text is perceived as less organized, or if this is an artifact (i.e., although subjects were not limited in the number of things they could list that they liked about the chapter, perhaps since they could not cite humor as something they liked about the chapter, they searched for an alternative explanation).

Dislikes

When asked what they disliked about each of the chapters, subjects listed a variety of things. About one-quarter of subjects (23 percent) said there was nothing they disliked about the experimental ideas chapter. In fact, readers of the nonhumorous experimental ideas chapter were somewhat more likely (14, compared to 10 of the readers of the humorous version) to say there was nothing they disliked. Readers of the nonhumorous versions of each chapter were more likely to say that those chapters were dry, boring, not interesting, or didn't hold their attention. One explanation for this effect is that, even if subjects didn't find the humor to be funny, the humor did provide novelty: it is unlikely that subjects encountered humor (even if the humor was a rather poor attempt at amusing the reader) in many of the textbooks they had read prior to their experiences in this study. According to Keller (1987), novelty can help to capture and maintain a learner's interest. Unfortunately, if humor works because of its novelty and not through some other mechanism, the novelty effect is likely to wear off after repeated exposure.

Aids to Learning

When asked what helped them learn the information in the chapters, about one-quarter of the readers of each of the humorous chapters (26 percent and 25 percent of the readers of the humorous experimental ideas chapter and the humorous scientific fairness chapter, respectively) said the humor helped them learn. Readers of both versions of both chapters frequently said that examples helped them learn; readers of the nonhumorous scientific fairness chapter cited examples somewhat more frequently than did readers of the humorous scientific fairness chapter. More readers of the humorous versions of the chapters said ease of reading or understanding, or the writing style helped them learn. In contrast, readers of the nonhumorous versions were more likely to say the overview or summary helped them learn. In other words, readers of the humorous versions were more likely to cite the process of the reading, while readers of the nonhumorous versions cited the structure.

Summary

The findings of this study support the assertion that humor can help students learn and can have other positive benefits (such as enhancing interest, enjoyment, and attention), but only under certain conditions. There were indications that the humor in the scientific fairness chapter was not well integrated with the text and therefore did not facilitate learning for readers. However, humor did not appear to have any negative effects, either on learning or on subjects' attitudes.

Regarding the qualitative data, it is difficult to know how subjects defined their uses of terms. For example, a subject's use of the term "interesting" when describing what he or she liked about a chapter or why he or she preferred it for leisure reading may have encompassed humor in addition to any number of other variables (e.g., relevance). Subjects also combined ideas in their responses that I classified into two or more categories. For example, when asked what helped to learn the information in the experimental ideas chapter, one subject wrote, "It was easy reading that grabbed my attention" (#106). This was classified under both "easy to read/understand" and "attention," but these factors may not have been separable in the subject's mind.

Based on my findings, I have revised my model showing the effect of humor in instructional text (see Figure 5-1). Learning, interest, and enjoyment remain as important outcomes of humor in instruction. Note that I have added subject knowledge and experiences and difficulty of content, which combine with actual humor in the text and delivery of the humor to influence subjects' perception and appreciation of humor. If the humor is perceived and appreciated by the subjects, then interest, enjoyment, and learning result. Although I didn't find that interest and enjoyment influenced learning in this study, I proposed that interest and enjoyment may result in a greater receptivity to additional experiences with the topic; more research is needed to determine if this occurs. Although there was a direct link between humor and learning for the experimental ideas chapter, none of the proposed mediating variables were found to have an effect on humor. However, there may be other variables that mediate the effect of humor on learning.

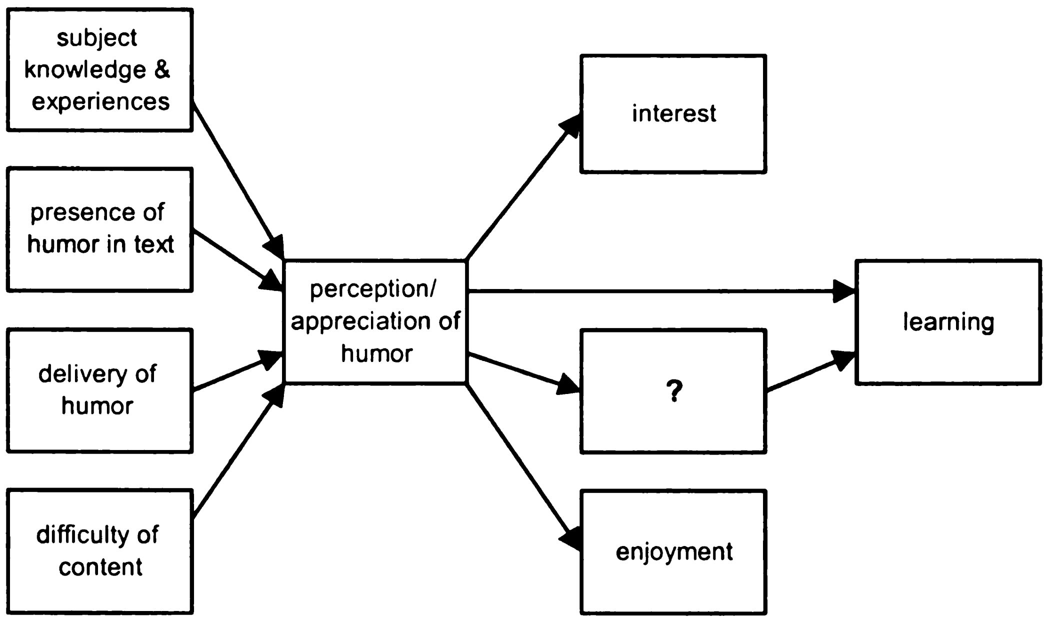

Figure 5-1. Revision of model, showing factors that influence the perception and appreciation of humor in instructional text and the outcomes of humor

Implications for Theory

In chapter 2, I cited dual coding theory as providing theoretical support for the use of humor in instructional text. Briefly, dual coding theory states that comprehension and memory of text are affected by both verbal and nonverbal systems (Sadoski, 93), and that we understand and remember more if both systems are engaged. I argued earlier that, based on students' reactions, the humor in the humorous scientific fairness was less integrated with the content. And although half of students who read the humorous version of the scientific fairness chapter said they liked the humor, only one quarter reported that humor helped them learn. If humor was indeed more effective in the experimental ideas chapter, then this research supports dual coding theory: the chapter that engaged both verbal and nonverbal systems was better remembered.

I also discussed Keller's ARCS model of motivational design of instruction earlier. Keller's ARCS model of motivation delineates four requirements that must be met in order for people to be motivated to learn: the student's **a**ttention must be gained and sustained, the material must be **r**elevant to the student, the student must be **c**onfident that he or she can learn the material, and the student must be **s**atisfied with the process of the experience (Keller, 1987). Keller's ARCS model includes two elements included in my original model: attention and relevance. I did not find a relationship between these variables and learning, but motivation by itself is not enough to cause learning. Other factors, such as the opportunity and incentive to learn, may override any effects of motivation.

Recommendations for Authors who Wish to Incorporate Humor into Instructional Text

Based on the results of my research (including this study, the pilot study, an early exploratory study of humor in learning, and reviews of the literature), I have the following recommendations for writers who wish to include humor in their written instructional materials:

- Make sure the humor is closely related to the text. Humor that is only peripherally related to the text may distract readers from the main ideas of the text.

- Make sure the humor does not assume background or experiences that students are unlikely to have. Students will not benefit from humor that they don't "get."
- Include humor only in instructional pieces that students are likely to understand reasonably easily. Humor does not appear to aid learning when readers are struggling just to make sense of the text.
- Include the humor as you are initially writing the material. Going back to the material later and attempting to add humor may result in humor that is more peripherally related to the ideas in the text and not as well integrated into the text as a whole.
- Link the humor to the main ideas of the text. Wade and Adams (1990) found that interesting ideas were recalled better than less interesting ideas that were more important.
- Build a relationship with the reader. In my initial studies of humor in instructional materials, I found that readers were taken aback by humor when there was no introduction by the author. When I included such an introduction (i.e., a note from the author telling his reasons for writing in the style he used), students were more receptive to the humor.
- Use humor with caution in particularly sensitive material. In the chapter on scientific fairness, after a discussion of a researcher who was investigated for research fraud and discredited, Martin wrote, "He then promptly committed suicide." Although this was not considered particularly humorous, at least one of the subjects felt this treated suicide lightly: when asked the question "What

did you not like about this chapter?", this subject wrote, "Suicide is not funny. It should not be used as a point of humor" (subject #195).

- Do not allow humor to take the place of effective instructional techniques in text. When asked what helped them learn the information in the chapter, subjects frequently cited organization, examples, and ease of reading.

These suggestions focus mostly on the circumstances under which it is appropriate to include humor in text, and not on how to come up with humor. Some professional organizations and companies that specialize in teaching public speaking offer seminars on including humor. In addition, there are books on comedy writing (e.g., Helitzer, 1987), and some community education programs offer courses in stand-up comedy and comedy writing. Authors who are interested in including humor in their textbooks may find it helpful to explore one or more of these opportunities.

Strengths

This study was a true experiment, which allowed me to manipulate the materials, determine subjects' reactions, and make cause-and-effect conclusions for the effects of humor.

Although I was not able to analyze the study as a single within-subjects design as I had intended, this design provided me with considerably more information than a between-subjects design would have. If I had designed a between-subjects study using only one of these chapters, I would have come to a different conclusion: had I done a study using only the experimental ideas chapter, I would have concluded that humor facilitates learning, while a study using only the scientific fairness chapter would have led to a conclusion that humor does *not* aid learning.

A power analysis performed prior to data collection indicated that 80 subjects would be sufficient to determine actual effects of humor. I ended up with usable data from slightly over 100 subjects, which I believe provided sufficient power to detect any true differences between the groups.

Limitations

A major limitation of this study was that it did not occur within the context of an actual course. Students were tested on recall without having either the ability or the incentive to study. Although the results I found may be important, my ability to generalize my findings to a classroom situation are extremely limited.

This study is limited to a single subject area (i.e., experimental psychology). I am not able to state the effects of humor in other subject areas, especially given that I discovered differences in the effects of humor between two chapters that came from a single book within a fairly specific subject area.

This study is also limited to college students who have been in school continuously since graduating from high school. Additional research would be needed to determine the effects of humor on subjects who are older and/or who have been out of school for a while.

Other than presence and absence of humor, I did not manipulate the amount of humor in this study to determine an optimal amount. Research has indicated that there is an optimal level of humor in a classroom situation (Ziv, 1988). However, spoken humor is qualitatively different from written humor: a presenter can rely on spontaneity and audience needs when presenting humor. These aspects are not available to a writer. Future studies may investigate the effects of differing levels of humor in text.

In order to look at how subjects' responses compare to specific instances of humor in the text, I had four people each rate the humor in two versions of text (either humorous experimental ideas and nonhumorous scientific fairness, or nonhumorous experimental ideas and humorous scientific fairness). Obviously, more raters would have helped me get a better sense of what was funny in each chapter and how the humor was interpreted differently by different people. However, practical considerations limited me to having each version of each chapter rated by two people.

Out of necessity, the text contained a visible author, which is very different from the majority of textbooks students have encountered. Some subjects may have found a visible author disconcerting or even unacceptable (Paxton, 1997). However, Paxton (1997) found positive effects of a visible author even for a student who found it objectionable.

Next Steps

Throughout this section, I have suggested additional research that is needed to confirm or further clarify my findings. In sum, additional studies should compare materials matched for type and amount of humor, general interest, and difficulty of content. Studies should take place in classrooms or other environments in which subjects have the incentive and opportunity to study. Research in an actual classroom setting, in which the students have a high motivation to do well, may result in the effects of humor washing out; however, if humor enhances interest and enjoyment, subjects may find learning to be easier. Subjects should be interviewed to determine what they find amusing in the chapters, and to find out other effects of humor (such as physiological arousal). Questions in tests of learning should link both to specific instances of humor in the text (to determine effects of specific instances of humor on learning) and to the text more generally (to verify the overall effect of humor in text). Additional studies should include populations other than college students and should include text from subjects other than experimental psychology in order to determine the more general effects of humor on learning.

REFERENCES

Agnes, M. (Ed.). (2000). *Webster's New World College Dictionary* (4 ed.). Foster City, CA: IDG Books Worldwide, Inc.

Albanese, M. A., & Mitchell, S. (1993). Problem-based learning: A review of literature on its outcomes and implementation issues. *Academic Medicine, 68*(1), 52-81.

Alexander, P. A., Kulikowich, J. M., & Jetton, T. L. (1994). The role of subject-matter knowledge and interest in the processing of linear and nonlinear texts. *Review of Educational Research, 64*(2), 201-252.

Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994). How subject-matter knowledge affects recall and interest. *Americal Educational Research Journal, 31*(2), 313-337.

Anandam, K., & Fleckman, B. (1977). Telephone dialogue intervention in open learning instruction. *Journal of College Student Personnel, 18*(3), 219-223.

Barreca, R. (1991). *They Used to Call Me Snow White... But I Drifted: Women's Strategic Use of Humor*. New York: Viking.

Berger, A. A. (1997). *The Art of Comedy Writing*. New Brunswick, NJ: Transaction Publishers.

Bergin, D. A. (1999). Influences on classroom interest. *Educational Psychologist, 34*(2), 87-98.

Bless, H. (2001). Mood and the use of general knowledge structures. In L. L. Martin & G. L. Clore (Eds.), *Theories of Mood and Cognition: A User's Handbook* (pp. 9-26). Mahwah, NJ: Lawrence Erlbaum Associates.

Bohlin, R. M., & Milheim, W. D. (1994). *Applications of an adult motivational instructional design model.* Paper presented at the Association for Educational Communications and Technology, Nashville, TN.

Brophy, J. (1998). *Motivating Students to Learn*. Boston: McGraw-Hill.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review, 64*, 123-152.

Bryant, J., Comisky, P. W., Crane, J. S., & Zillmann, D. (1980). Relationship between college teachers' use of humor in the classroom and students' evaluations of their teachers. *Journal of Educational Psychology, 72*(4), 511-519.

Bryant, J.S., & Zillmann, D. (1989). Using humor to promote learning in the classroom. In P.E. McGhee (Ed.), *Humor and Children's Development*. New York: Haworth.

Cabello, B., & Terrell, R. (1994). Making students feel like family: How teachers create warm and caring classroom climates. *Journal of Classroom Interaction, 29*, 17-23.

Cennamo, K., & Braunlich, E. (1996). *The effects of relevance on mental effort.* Paper presented at the National Convention of the Association for Educational Communications and Technology, Indianapolis, IN.

Clayden, G. S., & Wilson, B. (1988). Computer-assisted learning in medical education. *Medical Education, 22*, 456-467.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Cohen, T. (1999). *Jokes: Philosophical Thoughts on Joking Matters*. Chicago: The University of Chicago Press.

Collier, K. G. (1980). Peer-group learning in higher education: The development of higher order skills. *Studies in Higher Education, 5*(1), 55-61.

Crismore, A. (1984). The rhetoric of textbooks: Metadiscourse. *Journal of Curriculum Studies, 16*, 279-296.

Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory.* Fort Worth, TX: Harcourt Brace Jovanovich College Publishers.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York: Plenum.

Deci, E. L., & Ryan, R. M. (2000). Self-determination theory: An approach to human motivation and personality [On-line]. *Available: http:www.psych.rochester.edu/SDT/measures/intrins.html.*

Eisenberger, E., & Cameron, J. (1996). Detrimental effects of reward: Reality or myth? *American Psychologist, 51*, 1153-1166.

Elliott, E., & Dweck, C. S. (1988). Goals: An approach to motiation and achievement. *Journal of Personality and Social Psychology, 54*(5-12).

Fiedler, K. (2001). Affective states trigger processes of assimilation and accommodation. In L. L. Martin & G. L. Clore (Eds.), *Theories of Mood and Cognition: A User's Handbook* (pp. 85-98). Mahwah, NJ: Lawrence Erlbaum Associates.

Forgas, J. P. (1991). Affect and social judgments: An introductory review. In J. P. Forgas (Ed.), *Emotion and Social Judgments* (pp. 3-29). Elmsford, NY: Pergamon Press Inc.

Garner, R., & Gillingham, M. G. (1991). Topic knowledge, cognitive interest, and text recall: A microanalysis. *Journal of Experimental Education, 59*(4), 310-319.

Garrett, T. J., Ashford, A. R., & Savage, D. G. (1987). A comparison of computer-assisted instruction and tutorials in hematology and oncology. *Journal of Medical Education, 62*, 918-922.

Gathany, N. C., & Stehr-Green, J. K. (1994). Putting life into computer-based training: The creation of an epidemiologic case study. *Educational Technology*, 44-47.

Glew, P. (1983). Kinesiology: Academic aspects of physical education. In G. Collier (Ed.), *The Management of Peer-Group Learning: Syndicate Methods in Higher Education* (pp. 42-47). Guildford: Society for Research into Higher Education.

Glynn, S. M. (1978). Capturing readers' attention by means of typographical cuing strategies. *Educational Technology, 18*(11), 7-12.

Goetz, E. T., & Sadoski, M. (1995). Commentary: The perils of seduction: Distracting details or incomprehensible abstractions? *Reading Research Quarterly, 30*(3), 500-515.

Goldstein, J. (1993). Humor and comedy in mass media. *Medienpsychologie, 5*(4), 246-256.

Gorham, J., & Christophel, D. M. (1990). The relationship of teachers' use of humor in the classroom to immediacy and student learning. *Communication Education, 39*(1), 46-62.

Hartley, J. (1987). Designing electronic text: The role of print-based research. *Educational Communication and Technology Journal, 35*(1), 3-17.

Helitzer, M. (1987). *Comedy Writing Secrets: How to Think Funny, Write Funny, Act Funny and Get Paid for It*. Cincinnati, OH: Writers' Digest Books.

Hidi, S. (1990). Interest and its contribution as a mental resource for learning. *Review of Educational Research, 60*(4), 549-571.

Hidi, S. (2000). An interest researcher's perspective: The effects of intrinsic and extrinsic factors on motivation. In C. Sansone & J. Harackiewicz (Eds.), *Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance* (pp. 309-339). San Diego: Academic Press.

Isen, A. M., Daubman, K. A., & Gorgoglione, J. M. I. E., (pp.). :. (1987). The influence of positive affect on cognitive organization: Implications for education. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, Learning, and Instruction, Volume 3: Conative and Affective Process Analyses* (pp. 143-164). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Johnson, H. A. (1990). Humor as an innovative method for teaching sensitive topics. *Educational Gerontology, 16*, 547-559.

Kaplan, R. M., & Pascoe, G. C. (1977). Humorous lectures and humorous examples: Some effects upon comprehension and retention. *Journal of Educational Psychology, 69*(1), 61-65.

Karis, D., Fabiani, M., & Donchin, E. (1984). "P300" and memory: Individual differences in the von Restorff effect. *Cognitive Psychology, 16*, 177-216.

Keller, J. M. (1979). Motivation and instructional design: A theoretical perspective. *Journal of Instructional Development, 2*(4), 26-34.

Keller, J. M. (1983). Motivational design of instruction. In C. M. Reigeluth (Ed.), *Instructional-Design Theories and Models: An Overview of their Current Status* (pp. 383-434). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Keller, J. M. (1987). Strategies for stimulating the motivation to learn. *Performance and Instruction, 26*(8), 1-7.

Korobkin, D. (1988). Humor in the classroom: Considerations and strategies. *College Teaching, 36*(4), 154-158.

Krapp, A., Hidi, S., & Renninger, K. A. (1992). Interest, learning, and development. In K. A. Renninger & S. Hidi & A. Krapp (Eds.), *The role of interest in learning and development*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Long, D. L., & Graesser, A. C. (1988). Wit and humor in discourse processing. *Discourse Processes, 11*, 35-60.

Loomans, D., & Kolberg, K. (1993). *The Laughing Classroom: Everyone's Guide to Teaching With Humor and Play*. Tiburon, CA: HJ Kramer, Inc.

Lorch, R. F., Jr., Lorch, E. P., & Inman, W. E. (1993). Effects of signaling topic structure on text recall. *Journal of Educational Psychology, 85*(2), 281-290.

Lorch, R. F., Jr., Lorch, E. P., & Klusewitz, M. A. (1993). College students' conditional knowledge about reading. *Journal of Educational Psychology, 85*(2), 239-252.

Main, R. G. (1992). *Integrating the affective domain into the instructional design process* (AL-TP-1992-0004). Brooks Air Force Base, TX: California State Univ., Chico. Coll. of Communication.

Martin, D. W. (2000). *Doing Psychology Experiments* (5 ed.). Belmont, CA: Wadsworth/Thomson Learning.

Martin, R. A., & Lefcourt, H. M. (1983). Sense of humor as a moderator of the relation between stressors and moods. *Journal of Personality and Social Psychology, 45*(6), 1313-1324.

McDaniel, M. A., Waddill, P. J., Finstad, K., & Bourg, T. (2000). The effects of text-based interest on attention and recall. *Journal of Educational Psychology, 92*(3), 492-502.

Means, T. B., Jonassen, D. H., & Dwyer, F. M. (1997). Enhanced relevance: Embedded ARCS strategies vs. purpose. *Educational Technology Research and Development, 45*(1), 5-17.

Mogavero, D. T. (1979). It's confirmed: J-students like humor in the classroom. *Journalism Educator, 34*(1), 43-44, 52.

Narvaez, D., van den Broek, P., & Ruiz, A. B. (1999). The influence of reading purpose on inference generation and comprehension in reading. *Journal of Educational Psychology, 91*(3), 488-496.

Nolen, S. B. (1995). Effects of a visible author in statistical texts. *Journal of Educational Psychology, 87*(1), 47-65.

Okey, J. R., & Santiago, R. S. (1991). Integrating instructional and motivational design. *Performance Improvement Quarterly, 4*(2), 11-21.

Parrott, T. E. (1994). Humor as a teaching strategy. *Nurse Educator, 19*(3), 36-38.

Paxton, R. J. (1997). "Someone with like a life wrote it": The effects of a visible author on high school history students. *Journal of Educational Psychology, 89*(2), 235-250.

Peterson, I. (1980). Humor in the physics classroom. *The Physics Teacher, 18*(9), 646-650.

Robbins, J. (1994). Using humor to enhance learning in the skills laboratory. *Nurse Educator, 19*(3), 39-41.

Rowell, R. M., & Van Kirk, J. A. (1978). Tutorial modules: How we keep humans in our instructional packages. *Educational Technology, 18*(11), 13-15.

Ryan, R. M., Connell, J. P., & Plant, R. W. (1990). Emotions in nondirected text learning. *Learning and Individual Differences, 2*(1), 1-17.

Sadoski, M., Goetz, E. T., & Fritz, J. B. (1993). Impact of concreteness on comprehensibility, interest, and memory for text: Implications for dual coding theory and text design. *Journal of Educational Psychology, 85*(2), 291-304.

Sadoski, M., Goetz, E. T., & Rodriguez, M. (2000). Engaging texts: Effects of concreteness on comprehensibility, interest, and recall in four text types. *Journal of Educational Psychology, 92*(1), 85-95.

Sadoski, M., & Quast, Z. (1990). Reader response and long-term recall for journalistic text: The roles of imagery, affect, and importance. *Reading Research Quarterly, 25*(4), 256-272.

Schiefele, U., & Krapp, A. (1996). Topic interest and free recall of expository text. *Learning and Individual Differences, 8*(2), 141-160.

Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory and Cognition, 19*, 523-542.

Schmidt, S. R. (1994). Effects of humor on sentence memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 953-967.

Schmidt, S. R., & Williams, A. R. (2001). Memory for humorous cartoons. *Memory and Cognition, 29*(2), 305-311.

Schraw, G., & Lehman, S. (2001). Situational interest: A review of the literature and directions for future research. *Educational Psychology Review, 13*(1), 23-52.

Sheppard, L. M. (1998, April). *Transforming a basic computer literacy module from snoozer to success.* Paper presented at the Abstract presented at the meeting of the Central Group on Educational Affairs, a division of the Association of American Medical Colleges, Chicago, IL.

Spires, H. A., & Donley, J. (1998). Prior knowledge activation: Inducing engagement with informational texts. *Journal of Educational Psychology, 90*(2), 249-260.

Svebak, S. (1996). The development of the Sense of Humor Questionnaire: From SHQ to SHQ-6. *Humor: International Journal of Humor Research, 9*(3/4), 341-361.

Vance, C. M. (1987). A comparative study on the use of humor in the design of instruction. *Instructional Science, 16*, 79-100.

Veatch, T. C. (1998). A theory of humor. *Humor: International Journal of Humor Research, 11*(2), 161-215.

Wade, S. E., & Adams, R. B. (1990). Effects of importance and interest on recall of biographical text. *Journal of Reading Behavior, 22*(4), 331-353.

Wandersee, J. H. (1982). Humor as a teaching strategy. *The American Biology Teacher, 44*(4), 212-218.

Wanzer, M., Booth-Butterfield, M., & Booth-Butterfield, S. (1995). The funny people: A source-orientation to the communication of humor. *Communication Quarterly, 43*(2), 142-154.

Watson, M. J., & Emerson, S. (1988). Facilitate learning with humor. *Journal of Nursing Education, 27*(2), 89-90.

Weinberger, M. G., & Gulas, C. S. (1992). The impact of humor in advertising: A review. *Journal of Advertising, 21*(4), 35-59.

Welkowitz, J., Ewen, R. B., & Cohen, J. (2000). *Introductory Statistics for the Behavioral Sciences* (5 ed. Vol. 5). Fort Worth, TX: Harcourt Brace & Company.

Zemke, R. (1991). Humor in training: Laugh and the world learns with you--maybe. *Training, 28*(8), 26-29.

Zillmann, D. (1988). Mood management through communication choices. *American Behavioral Scientist, 31*(3), 327-340.

Zillmann, D., & Bryant, J. (1985). Affect, mood, and emotion as determinants of selective exposure." In D. Zillmann & J. Bryant (Eds.), *Selective Exposure to Communication* (pp. 157-190). HIllsdale, NH: Lawrence Erlbaum.

Ziv, A. (1988). Teaching and learning with humor: Experiment and replication. *Journal of Experimental Education, 57*(1), 5-15.

APPENDICES

APPENDIX A

How to Get an Experimental Idea

Adapted from *Doing Psychology Experiments, 5th edition,*
by D. W. Martin. ©2000. Reprinted with permission of Wadsworth, an imprint of the Wadsworth Group, a division of Thomson Learning.
Fax 800 730-2215.

Author's note: I have attempted to write Doing Psychology Experiments *in a friendly style. While those who teach psychology have finally admitted that introductory texts need not be written in a pedantic style, we still seem to be unwilling to write about more interesting material in an interesting way. Hopefully this text will be a forerunner of a number of less formal advanced texts.*

I once was so bold and foolish as to assign introductory psychology students the task of proposing seven experiments as a course requirement. At first I was puzzled by their reactions to this assignment. Above the din of gnashing teeth, the moaning, and the groaning could be heard the plaintive wail of my stupefied students, "How do we get an idea?" Not only did I find it difficult to understand why getting an idea would pose such a problem, I also found it impossible to answer the question. I have now pondered this pervasive problem and formed an opinion about why it occurs and what can be done to solve it.

I don't believe the problem is that students have no ideas. As small children, we are curious about everything, including human behavior: "Mommy, why is that man so fat?" "How does Jenny eat with her left hand?" "Why can't I spell as good as Betty?" "Why is Donald so weird?" I refuse to believe that this curiosity simply fades away. In fact, the same students who "could not get an idea" seem to have plenty of thoughts about human behavior at parties or bull sessions: "What's the best way to study for my bio exam?" "Should I marry him or just move in with him?" "Am I more creative on days that I eat broccoli first thing in the morning?"

For this reason I will refuse to believe you if you tell me you don't have any ideas for an experiment. It's not true that you don't have any ideas, but it may be true that you are afraid something is wrong with the ideas you do have. This fear can paralyze your natural creativity, and, after a while, all your ideas seem inadequate to you.

Fearing Experimental Ideas

Fears about experimental ideas are usually irrational, stemming from a misunderstanding of psychology experiments. Psychologists call irrational fears phobias. Since I am a psychologist, I cannot resist the temptation to name the phobias behind our inability to get experimental ideas. (Any resemblance of these names to accepted psychological terminology is purely coincidental.) The following phobias seem to be the most common.

Geniephobia (Fear of Geniuses). Geniephobia stems from the common belief that anyone doing research must be a genius and that your modest brainpower couldn't possibly measure up. Researchers often do little to counteract this belief, and a few have been known to cultivate it. For years, every time I read a journal article I pictured the author as a wise-looking old man with flowing white locks. It was a shock to find that many experimenters are young, ordinary-looking people who make silly mistakes and say stupid things just like the rest of us.

My own geniephobia is still being cured. The more experimental psychologists I meet, the less I think only geniuses can do this kind of work. (I do not mean to imply that experimental psychologists are dumber than other scientists. Biologists and physicists can be dumb, too.) So, relax. Your ideas are probably as good as theirs were when they were getting started.

Imitatophobia (Fear of Imitating). Those people with imitatophobia are afraid to propose any idea unless it is absolutely original. An imitatophobic who combines this fear with a belief that everything worthwhile has already been thought of often reaches a state of total paralysis. Truly original experiments are few indeed in psychology. Most experiments use variations of somebody else's method to test somebody else's theory. In the next chapter you will learn how to find out what other experiments have been done in your area of interest, and you will find out exactly how unoriginal you are. Don't be afraid to move science along in small steps, however. That's what the rest of us do.

Paraphernaliophobia (Fear of Apparatus) and Manuphobia (Fear of Doing It by Hand). If the sum total of your mechanical knowledge of the automobile is that the right pedal makes it go and the left pedal makes it stop, you are a prime candidate for paraphernaliophobia. This malady will scare you away from any experimental idea requiring apparatus more sophisticated than a deck of cards.

On the other hand, if you will not consider doing any experiment unless it requires sophisticated scientific equipment, you are a victim of the opposite affliction, manuphobia. Everyone knows that the more complex the equipment, the better the research.

Both positions are unfounded, however. Some of the best research uses little or no apparatus. Piaget developed a major area of child psychology with no more apparatus than toy blocks, water glasses, and modeling clay. Other areas of psychology, such as verbal learning, concept formation, attitude assessment, and personality, require no more than pencil and paper. Apparatus can help you do research, but it isn't the research itself. And when apparatus is necessary, someone will be available who can teach you how to use it.

Parsimoniophobia (Fear of Simplicity). Parsimoniophobics think they must come up with grandiose experiments that will change the course of science in one fell swoop. Their motto is: If it's simple, it can't be science. While there are some advantages to complex experiments, generally you should aim for the simplest experiment that can answer your experimental question. People with parsimoniophobia seldom complete their majestic experiments; when they do, they usually cannot interpret the results. To start with, then, think simple. You can always pursue more complex questions later.

Calculatophobia (Fear of Statistics). Some people dread having to do any calculations tougher than counting their fingers. If you never can remember how to figure out your car's gas mileage or how to keep your checkbook straight, you are a potential calculatophobic. If you will consider only those experiments that require the simplest statistical test, remember that such tests are tools that can help you interpret your results; they should not cause you to throw out good experimental ideas. You can always find someone who enjoys playing with numbers to help you analyze your data. I am not saying that a knowledge of statistics is unimportant, but it is, after all, just a tool used in science, not science itself.

Imperfectaphobia (Fear of Being Imperfect). An imperfectaphobic will not tell you about an experimental idea until every tiny detail is perfectly worked out, and his or her proposal looks like the final report. This attitude often stems from having read too many pristinely presented journal articles. But keep in mind that journal articles are end products; they seldom reflect the sloppy thinking and general air of confusion that precedes most experiments. Completed experiments are often quite different from the experimenters' original ideas. The original idea for an experiment simply forms the kernel; the experimental procedure will evolve as you set up and conduct the experiment. If you take the plunge and begin talking about your experiment in rough form, others may be able to help you mold it into a perfect experiment. Well, almost perfect.

Pseudononphonoscientiaphobia (fear of not sounding scientific). People with this hideous affliction can only recognize a great idea if it is expressed in **scientific jargonese.**[1] Scientific jargonese is a pseudolanguage that some scientists make up to sound good when they talk with other scientists who do similar research. It helps obscure the research from the general public and sometimes from other scientists as well. For example, in jargonese an experiment designed to determine if people remember words better when the words are in groups is described as an investigation into "the effect of taxonomic and categorical clustering on the retention of verbal material." Or a notion that people from ethnic groups live in the same neighborhood because of pressure from their friends is described as an experiment examining "the effect of demographic distribution by ethnic affiliation as a function of peer-group coercion." Jargonese can also be translated into everyday language. If you are interested in "the effect of affiliative preference on the salience of dimensions in person perception," you are actually trying to find out whether people who join certain organizations differ in the way they see other people. Try translating one yourself: "the effects of maternal employment on sibling aggressive tendencies." If you came close to "Do working mothers' kids fight more?" then you are catching on. Be sure to buy my next book: *Scientific Jargonese for Fun and Profit.*

As you can guess from the way this book is written, I believe that most pseudoscientific jargonese is nonsense. Good scientists need not hide behind their language. A good idea is a good idea regardless of the words used to express it.

Ergophobia (fear of work). Sorry, there is no known cure for this affliction.

Now that we are aware of what fears might block our creativity, let's try to get some experimental ideas. What is the best way to start?

Direct Observation

Someone once said that it's easy to write: just sit at the typewriter and stare at the keyboard until drops of blood appear on your forehead. This also describes the best way to avoid coming up with experimental ideas. Since we are interested in human behavior rather than typewriter behavior, the best thing to do is observe humans, not typewriters!

[1] I am using the word **jargonese** to represent the fourth dictionary definition of jargon (speech or writing characterized by pretensions, terminology, and involved syntax) as opposed to the first definition (the language peculiar to a particular trade, profession, or group). The line between jargon and jargonese is thin indeed.

Getting experimental ideas is simply a matter of noticing what goes on around you. Once you become a good observer, your natural curiosity will provide you with experimentally testable questions. One week of constant observation should provide you with enough experiments to last three careers.

Indeed, some of the classic research in experimental psychology started with a simple observation. If Ekhard Hess's wife had not noticed that his pupils got bigger when he was looking at bird pictures, pupillometrics might never have gotten underway. If Ivan Pavlov had not noticed that his dogs were salivating to stimuli other than meat powder, Igor Nosnoranovitch might have been the father of classical conditioning instead. If Jean Piaget had not noticed that his daughter Jacqueline stopped making gurgling noises when she could no longer see her bottle, he might have been a famous Swiss watchmaker. Most revolutionary experimental ideas have been generated by simple observation.

Public observation. After reading the next couple of paragraphs, take a paper and pencil, leave the room you are in, and walk outside where there are people to observe. As a training exercise in observation, make notes of possible experimental questions that occur to you as you stroll around.

First I'll go on a stroll to show you what I mean: I wander outside, and I see that the sun is shining.

1. Do people get more or less work done when the weather is nice?

I walk past two workmen laying concrete for a bike rack. One is working; the other is standing and watching.

2. Do workers stand around more when they are unionized?

A couple of joggers run by.

3. Do people who exercise regularly sleep better at night?

A young woman is sitting over there under a tree with a young bearded fellow. They are looking rather amorous, and I feel like a peeping Tom. Better move on.

4. Do women find men with beards more attractive than men without beards?

I see a large group of students filing into a classroom.

5. Do students in large lecture classes make better grades than those in small classes?

I arrive at a crosswalk. Will that car stop? Yes, it did. Across I go.

6. Are drivers more likely to give the right of way to pedestrians of the opposite sex?

I stop to watch a sports car zooming down the street.

7. Do people drive sports cars faster than regular cars?

I head back past the library.

8. Do students who study in the library retain information better than those who study in the dorm?

I pass the bike rack at the front of my office building and see lots of bikes.

9. Are ten-speed bikes easier to ride, than three-speed bikes?

I lope upstairs to my office. I am back.

I just got nine potential ideas for experiments. That's almost one per minute! Now you try it while I wait. Look around you. Make and observation and form a question.

Welcome back. Did you get any ideas? Think about the ideas you could get if you were that observant all the time. Now the problem is "which idea should I turn into an experiment?" because not all important questions can be answered by experiment.

For example, science cannot answer moral questions, such_as: "Is abortion wrong?" "Is it proper for women to wear short skirts?" "Is dope evil?" While we can certainly use the scientific method to determine people's opinions about these questions, we cannot devise any test that could answer the questions themselves. We must therefore eliminate all such questions from any list of experimental ideas. Other questions fail because they are not observable: "Do dogs think like humans?" "Is my experience of the color red the same as yours?" Finally, some questions fail experimentally because they cannot be reliably repeated. Some supporters of ESP (extrasensory perception), for example, claim that ESP occurs only under certain

conditions and that it is impossible to predict when the conditions are right. In other words, ESP works only some of the time. As long as this basic tenet governs ESP effects, it is impossible to test for the existence of ESP

Do all the questions in your list of ideas meet the ROT (i.e., repeatable, observable, testable) requirements? Take a moment to go through your list and eliminate any that fail to do so.

Some questions must be answered by correlational observation rather than by experimentation. For example, if as in question 7 we want to know whether people who choose to drive sports cars drive at a faster speed than those who drive other types of cars, we must do a correlational observation to answer the question. On the other hand, if we wish to know whether any driver tends to drive faster when driving a sports car, we could design an experiment to answer the question. Take another look at your list of ideas, and label each idea experimental or correlational.

Our little walk has been interesting, but people in public provide us with a limited set of behaviors. Whom else can we observe?

Observing yourself. As you may know, introspection was one of the earliest techniques in experimental psychology. Introspectionists, however, concentrated on looking at their own mental processes rather than their own behavior. Since a controversy developed about whether a person can know his or her own mental processes, experimental psychologists stopped watching themselves altogether. Rather than follow the dictum "Know thyself," they resolved to "Know not thyself." (Some experimental psychologists still don't know who they are.) It is still generally frowned on to do an experiment with yourself as the only subject; nevertheless, you can get some good experimental ideas this way. Not only will you be able to collect many samples of the behavior you are interested in, but you might even have some idea why you did what you did. The former can give you an idea for an experiment, the latter an idea for a theory.

With a little effort, you can begin to notice your own behavior. It may seem ridiculous to suggest that you do not notice yourself, but it is probably true. When dressing; which arm do you put into your shirt or blouse first? When you brush your teeth, do you brush the left side first or the right? Do you put the key to your house or room into the lock rightside up or upside down? When you cross your legs, do you put your left leg or your right leg on top more often? These are all things you do every day. Do you notice them? Observing yourself can be entertaining[2] as well as a good source of ideas. Write down the ideas as they occur to you.

Observing your friends. Your friends are also good sources of experimental ideas. It is important, however, to observe their behavior in as unobtrusive a manner as possible. Staring is considered impolite at best and grounds for a fight at worst. People sometimes avoid paying attention to their own behavior because they are not particularly fond of it themselves. Consequently, to avoid losing friends, keep your observations to yourself. Pointing out your insights, no matter how brilliant, will not help you win friends and influence people.

[2] If you develop this skill, you will have to learn to control yourself in public. You may be considered strange if you break into gales of laughter over your own behavior.

Observing children. Observing children is a necessity if you are interested in doing experiments in the area of developmental psychology, but children can also give you good ideas for other areas of research. If you are not blessed with (or plagued by, depending on your point of view) children, you probably have friends and relatives who would be more than happy to let you watch theirs for a while. Unlike adults, who have learned that their behavior should appear rational, logical, and consistent to an outside observer, children generally behave in ways that are uncomplicated by complex patterns or social inhibitions. Because most kids couldn't care less about adult standards, you will be able to observe relatively uncontaminated behavior patterns in children.

Observing pets. Animals are interesting to study in their own right, but much of their behavior can also be generalized to humans.

Furthermore, you will find that pets are even less inhibited than children; since they are less capable of highly complex behavior patterns, their behavior is often easier to interpret. In addition, you can manipulate your pet's environment without worrying as much about the moral implications of possible permanent damage.

Vicarious Observation

Although you may find it less exciting than direct observation, you can also get experimental ideas, by reading other people's research. You might feel that this technique of **vicarious observation** feeds off people's creativity, but nonetheless this approach has certain practical advantages. For one thing, because the broad experimental question you are researching already has a stamp of approval from the author and journal reviewers, you know that the questions being asked are considered important. Second, somebody else has already fit the experimental result into the existing body of knowledge, thereby structuring the area of research for you and saving you time and effort. Finally, earlier researchers have devised a method of attack that apparently works and that you may be able to modify and use in your research.

In beginning your search for an idea, you should first identify an area of research that interests you. You will then know what types of journals you should read. Your topic should be as specific as possible: competition in small groups, play therapy in schools, perception of visual illusions, development of arithmetic abilities, and so on. For the more general topics you specify, you can simply scan journals having related articles. For more specific topics, this procedure is rather inefficient, and you will need to do a literature search. In either case, as you read the journal articles, make a note of the experimental questions left unanswered by the research. The author will sometimes help you discover what these questions are by suggesting where future research should go. You might select one of these experimental questions for your experiment.

Expanding on Your Own Research

Once you have done several experiments, you will find that your own research may provide many experimental ideas. Every experiment you do will leave a number of questions unanswered. For example, after using several levels of an independent variable in an experiment, you may want to see what happens when you choose other levels. Or you may have controlled a certain variable at a particular level in one experiment and may wonder what would happen if you set it at a different level. Or you may come up with unexpected results and want to find out why the outcome was not as predicted. Each experiment usually brings up more unsolved questions than it answers.

This picture of science as a continual growth of new questions is different from that held by many people who think of science as a fixed body of knowledge we need only uncover. This latter concept views scientific research as leaving fewer and fewer questions unanswered as it proceeds. In reality, however, each experiment actually increases the number of questions to be answered. Instead of working ourselves out of business, we are working ourselves into more business than we can possibly handle.

This open-ended view of science can be very discouraging and very exciting. It can be discouraging because it is sometimes difficult to chart our progress through an ever-expanding universe in which we sometimes seem to take five steps backward for every step forward. On the other hand, it is exciting because we end up asking better and better questions. Perhaps the goal of science is not to find answers to all possible experimental questions but to answer ever more promising and important questions. In following up on your own research, you will find that your main problem is not "How can I get an experimental idea?" but "Which idea is the most important one to work on?"

Focus on a Practical Problem

If you have a practical problem crying for a solution, you may also have the kernel of an experimental idea. So far in this chapter we have been pretending that all research is **basic research**[3]—that is, research done for the sole purpose of increasing the scientific body of knowledge. It is not necessary to justify such research as a solution to any practical problem, although basic research can help solve practical problems. Today's behavior-modification techniques, which provide some of the most powerful procedures for correcting human behavior problems, are based on basic research done in the rat laboratories of yesteryear. Adams (1972) found that many of our military systems were designed using information from basic research done more than 20 years earlier. Thus, basic research does at times prove valuable for solving practical problems. In general, however, such research is done in the name of the advancement of science.

Research designed to solve a specific practical problem is called **applied research**. Perhaps you need to know how humans read handwritten numbers so that you can design a machine to automatically read ZIP codes. Or you may wish to know whether daily quizzes will improve a student's classroom performance on major examinations. Or you may want to know whether transactional analysis is a more effective therapy than psychoanalysis. Many such practical problems need immediate answers that basic researchers might never find. Needing an answer to a practical problem is a perfectly legitimate reason for doing research, and it can be satisfying if your findings have an immediate impact on the world.

Observation is again the key to getting ideas for applied research. Finding a practical problem is simply a matter of carefully observing human behavior and permitting your curiosity free rein. As with the other procedures that we have discussed for getting experimental ideas, you will find that more practical problems need to be solved than you can possibly do experiments to solve. As before, the question becomes "What should I do first?" not "What can I do?"

Summary

Although we all have natural curiosity about human behavior, many of us develop irrational fears that block our ideas. Some of us fear that all other researchers are geniuses and that our ideas will not be original. Some people are afraid to propose an experiment requiring complex apparatus, while others are afraid of experiments with simple apparatus. Others fear that their idea is too simple, that their experiment will require complicated statistics, or that their idea is not perfect when it is proposed. Finally, many people do not believe they have good ideas until they translate them into scientific jargonese.

The major key to getting experimental ideas is to learn to **observe** the world about you. You also need to know which ideas are scientifically appropriate. Ideas must be **repeatable, observable,** and **testable** to be experimental ideas. To get ideas you can observe yourself, friends, children, and even pets. Although the best ideas come from direct observation, you can also get ideas by reading other people's research (vicarious observation) and from following up your own research.

[3] Basic research is sometimes also called **pure research**, perhaps because one is not supposed to have mixed motives for doing it. Unfortunately, some people who do this kind of research seem to prefer other dictionary definitions of pure—for example, untainted with evil or guilt. I have never heard pure scientists defend the position that they are physically chaste, even though I suspect that this is the subconscious reason behind wearing white lab coats.

Finally, some of the best ideas come from having a practical problem that needs to be solved. Research designed to solve a practical problem is called **applied research,** while research aimed primarily at advancing scientific knowledge is called **basic** or **pure research.**

References

Adams, J. A. (1972). Research and the future of engineering psychology. *American Psychologist,* 27, 615-622.

How to Get an Experimental Idea

Adapted from *Doing Psychology Experiments, 5th edition,*
by D. W. Martin. ©2000. Reprinted with permission of Wadsworth,
an imprint of the Wadsworth Group, a division of Thomson Learning.
Fax 800 730-2215.

Author's note: I have attempted to write Doing Psychology Experiments *in a friendly style. While those who teach psychology have finally admitted that introductory texts need not be written in a pedantic style, we still seem to be unwilling to write about more interesting material in an interesting way. Hopefully this text will be a forerunner of a number of less formal advanced texts.*

I once assigned introductory psychology students the task of proposing seven experiments as a course requirement. At first I was puzzled by their reactions to this assignment. After a fair amount of complaining, they asked, "How do we get an idea?" Not only did I find it difficult to understand why getting an idea would pose such a problem, I also found it impossible to answer the question. I've thought a lot about this problem, and I have since formed an opinion about why it occurs and what can be done to solve it.

I don't believe the problem is that students have no ideas. As small children, we are curious about everything, including human behavior: "Mommy, why is that man so tall?" "How does Jenny eat with her left hand?" "Why can't I spell as good as Betty?" "Why do Tommy's parents spank him so much?" I refuse to believe that this curiosity simply fades away. In fact, the same students who "could not get an idea" seem to have plenty of thoughts about human behavior: "What's the best way to study for my bio exam?" "Should I marry him or just move in with him?" "Am I more creative in the morning?"

For this reason I will refuse to believe you if you tell me you don't have any ideas for an experiment. It's not true that you don't have any ideas, but it may be true that you are afraid something is wrong with the ideas you do have. This fear can paralyze your natural creativity, and, after a while, all your ideas seem inadequate to you.

Barriers to Developing Experimental Ideas

Fears about experimental ideas are usually irrational, stemming from a misunderstanding of psychology experiments. Below are some of the most common barriers students encounter when attempting to come up with ideas for experiments.

Lack of confidence. This stems from the common belief that anyone doing research must be very bright and that you couldn't possibly measure up. For years, whenever I read a journal article, I pictured the author as a wise and distinguished-looking older man. In reality, many experimenters are young, ordinary-looking people just like the rest of us.

My own fears in this area are still being addressed. But the more experimental psychologists I meet, the more I realize that they are regular people. Your ideas are probably as good as theirs were when they were getting started.

Unrealistic expectations. Many students are afraid to propose any idea unless it is absolutely original. If you combine this fear with a belief that everything worthwhile has already been thought of, you'll often be unable to come up with ideas. Truly original experiments are few indeed in psychology. Most experiments use variations of existing methods to test existing theories. In the next chapter you will learn how to find out what other experiments have been done in your area of interest. Don't be afraid to move science along in small steps.

Apparatus. If you feel you lack basic mechanical knowledge, you may be tempted to avoid any experiment that requires the use of apparatus. On the other hand, if you will not consider doing any experiment unless it requires sophisticated scientific equipment, you may (incorrectly) believe that the use of complex equipment leads to better research.

Both positions are unfounded, however. Some of the best research uses little or no apparatus. Piaget developed a major area of child psychology with no more apparatus than toy blocks, water glasses, and modeling clay. Other areas of psychology, such as verbal learning, concept formation, attitude assessment, and personality, require no more than pencil and paper. Apparatus can help you do research, but it isn't the research itself. And when apparatus is necessary, someone will be available who can teach you how to use it.

Complexity. Some people think they must come up with experiments that will change the course of science. While there are some advantages to complex experiments, generally you should aim for the simplest experiment that can answer your experimental question. People who tend toward complex experiments seldom complete them; when they do, they usually have difficulty interpreting the results. To start with, then, think simple. You can always pursue more complex questions later.

Statistics. Some people dread having to do any moderately complex calculations. If you have trouble calculating your car's gas mileage or balancing your checkbook, you may find statistics to be a significant barrier to your ability to develop ideas. If you will consider only those experiments that require the simplest statistical test, remember that such tests are tools that can help you interpret your results; they should not cause you to throw out good experimental ideas. You can always find someone to help you analyze your data. I am not saying that a knowledge of statistics is unimportant, but it is, after all, just a tool used in science, not science itself.

Perfection. Some people will not tell you about an experimental idea until every detail is perfectly worked out and the proposal looks like the final report. This attitude often stems from having read too many journal articles. Keep in mind that journal articles are end products; they do not reflect the enormous amount of thought and investigation that precedes most experiments.

Completed experiments are often quite different from the experimenters' original ideas. The original idea for an experiment simply forms the kernel; the experimental procedure will evolve as you set up and conduct the experiment. If you begin talking about your experiment in rough form, others may be able to help you mold it into a perfect experiment.

Sounding scientific. Some people can only recognize a great idea if it is expressed in **scientific jargonese.**[1] Scientific jargonese is a

[1] I am using the word **jargonese** to represent the fourth dictionary definition of jargon (speech or writing characterized by pretensions, terminology, and involved syntax) as opposed to the first definition (the language peculiar to a particular trade, profession, or group). The line between jargon and jargonese is thin indeed.

pseudolanguage that some scientists make up to sound good when they talk with other scientists who do similar research. It helps obscure the research from the general public and sometimes from other scientists as well. For example, in jargonese an experiment designed to determine if people remember words better when the words are in groups is described as an investigation into "the effect of taxonomic and categorical clustering on the retention of verbal material." Or a notion that people from ethnic groups live in the same neighborhood because of pressure from their friends is described as an experiment examining "the effect of demographic distribution by ethnic affiliation as a function of peer-group coercion." Jargonese can also be translated into everyday language. If you are interested in "the effect of affiliative preference on the salience of dimensions in person perception," you are actually trying to find out whether people who join certain organizations differ in the way they see other people. Try translating one yourself: "the effects of maternal employment on sibling aggressive tendencies."[2]

As you can guess from the way this book is written, I believe that most pseudoscientific jargonese is nonsense. Good scientists need not hide behind their language. A good idea is a good idea regardless of the words used to express it.

Work. Unfortunately, work is unavoidable when conducting an experiment.

Now that we are aware of what fears might block our creativity, let's try to get some experimental ideas. What is the best way to start?

[2] If you came close to "Do working mothers' kids fight more?" then you are catching on.

Direct Observation

Getting experimental ideas is simply a matter of noticing what goes on around you. Once you become a good observer, your natural curiosity will provide you with experimentally testable questions. One week of constant observation should provide you with enough experiments to last three careers.

Indeed, some of the classic research in experimental psychology started with a simple observation. If Ekhard Hess's wife had not noticed that his pupils got bigger when he was looking at bird pictures, pupillometrics might never have gotten underway. If Ivan Pavlov had not noticed that his dogs were salivating to stimuli other than meat powder, someone else might have been the father of classical conditioning instead. If Jean Piaget had not noticed that his daughter Jacqueline stopped making gurgling noises when she could no longer see her bottle, he might not be known for his theories of development. Most revolutionary experimental ideas have been generated by simple observation.

Public observation. After reading the next couple of paragraphs, take a paper and pencil, leave the room you are in, and walk outside where there are people to observe. As a training exercise in observation, make notes of possible experimental questions that occur to you as you stroll around.

First I'll go on a stroll to show you what I mean: I wander outside, and I see that the sun is shining.

1. Do people get more or less work done when the weather is nice?

I walk past two workmen laying concrete for a bike rack. One is working; the other is standing and watching.

2. Do workers stand around more when they are unionized?

A couple of joggers run by.

3. Do people who exercise regularly sleep better at night?

A young woman is sitting over there under a tree with a young bearded fellow. They are looking rather amorous.

4. Do women find men with beards more attractive than men without beards?

I see a large group of students filing into a classroom.

5. Do students in large lecture classes make better grades than those in small classes?

I arrive at a crosswalk. Will that car stop? Yes, it did. Across I go.

6. Are drivers more likely to give the right of way to pedestrians of the opposite sex?

I stop to watch a sports car zooming down the street.

7. Do people drive sports cars faster than regular cars?

I head back past the library.

8. Do students who study in the library retain information better than those who study in the dorm?

I pass the bike rack at the front of my office building and see lots of bikes.

9. Are ten-speed bikes easier to ride than three-speed bikes?

I lope upstairs to my office. I am back.

I just got nine potential ideas for experiments. That's almost one per minute! Now you try it while I wait. Look around you. Make an observation and form a question.

Welcome back. Did you get any ideas? Think about the ideas you could get if you were that observant all the time. Now the problem is "which idea should I turn into an experiment?" because not all important questions can be answered by experiment.

For example, science cannot answer moral questions, such as: "Is abortion wrong?" "Are drugs evil?" While we can certainly use the scientific method to determine people's opinions about these questions, we cannot devise any test that could answer the questions themselves. We must therefore eliminate all such questions from any list of experimental ideas. Other questions fail because they are not observable: "Do dogs think like humans?" "Is my experience of the color red the same as yours?" Finally, some questions fail experimentally because they cannot be reliably repeated. Some supporters of ESP (extrasensory perception), for example, claim that ESP occurs only under certain conditions and that it is impossible to predict when the conditions are right. In other words, ESP works only some of the time. As long as this basic tenet governs ESP effects, it is impossible to test for the existence of ESP

Do all the questions in your list of ideas meet the repeatable, observable, and testable requirements? Take a moment to go through your list and eliminate any that fail to do so.

Some questions must be answered by correlational observation rather than by experimentation. For example, if as in question 7 we want to know whether people who choose to drive sports cars drive at a faster speed than those who drive other types of cars, we must do a correlational observation to answer the question. On the other hand, if we wish to know whether any driver tends to drive faster when driving a sports car, we could design an experiment to answer the question. Take another look at your list of ideas, and label each idea experimental or correlational.

Our little walk has been interesting, but people in public provide us with a limited set of behaviors. Whom else can we observe?

Observing yourself. As you may know, introspection was one of the earliest techniques in experimental psychology. Introspectionists, however, concentrated on looking at their own mental processes rather than their own behavior. Since a controversy developed about whether a person can know his or her own mental processes, experimental psychologists stopped watching themselves altogether. It is still generally frowned on to do an experiment with yourself as the only subject; nevertheless, you can get some good experimental ideas this way. Not only will you be able to collect many samples of the behavior you are interested in, but you might even have some idea why you did what you did. The former can give you an idea for an experiment, the latter an idea for a theory.

With a little effort, you can begin to notice your own behavior. It may seem ridiculous to suggest that you do not notice yourself, but it is probably true. When dressing; which arm do you put into your shirt or blouse first? When you brush your teeth, do you brush the left side first or the right? Do you put the key to your house or room into the lock rightside up or upside down? When you cross your legs, do you put your left leg or your right leg on top more often? These are all things you do every day. Do you notice them? Observing yourself can be a good source of ideas. Write down the ideas as they occur to you.

Observing your friends. Your friends are also good sources of experimental ideas. It is important, however, to observe their behavior in as unobtrusive a manner as possible. Most people, if they know they are being observed, will change their behavior. For the same reason, it's also important to keep your observations to yourself. Pointing out your insights is likely to make those you are observing self-conscious.

Observing children. Observing children is a necessity if you are interested in doing experiments in the area of developmental psychology, but children can also give you good ideas for other areas of research. If you are not blessed with children, you probably have friends and relatives who would be more than happy to let you watch theirs for a while. Unlike adults, who have learned that their behavior should appear rational, logical, and consistent to an outside observer, children generally behave in ways that are uncomplicated by complex patterns or social inhibitions. Because most kids couldn't care less about adult standards, you will be able to observe relatively uncontaminated behavior patterns in children.

Observing pets. Animals are interesting to study in their own right, but much of their behavior can also be generalized to humans. Furthermore, you will find that pets are even less inhibited than children; since they are less capable of highly complex behavior patterns, their behavior is often easier to interpret. In addition, you can manipulate your pet's environment without worrying as much about the moral implications of possible permanent damage.

Vicarious Observation

Although you may find it less exciting than direct observation, you can also get experimental ideas, by reading other people's research. You might feel that this technique of **vicarious observation** feeds off people's creativity, but nonetheless this approach has certain practical advantages. For one thing, because the broad experimental question you are researching already has a stamp of approval from the author and journal reviewers, you know that the questions being asked are considered important. Second, somebody else has already fit the experimental result into the existing body of knowledge, thereby structuring the area of research for you and saving you time and effort. Finally, earlier researchers have devised a method of approaching the issue that works and that you may be able to modify and use in your research.

In beginning your search for an idea, you should first identify an area of research that interests you. You will then know what types of journals you should read.

Your topic should be as specific as possible: competition in small groups, play therapy in schools, perception of visual illusions, development of arithmetic abilities, and so on. For the more general topics you specify, you can simply scan journals having related articles. For more specific topics, this procedure is rather inefficient, and you will need to do a literature search. In either case, as you read the journal articles, make a note of the experimental questions left unanswered by the research. The author will sometimes help you discover what these questions are by suggesting where future research should go. You might select one of these experimental questions for your experiment.

Expanding on Your Own Research

Once you have done several experiments, you will find that your own research may provide many experimental ideas. Every experiment you do will leave a number of questions unanswered. For example, after using several levels of an independent variable in an experiment, you may want to see what happens when you choose other levels. Or you may have controlled a certain variable at a particular level in one experiment and may wonder what would happen if you set it at a different level. Or you may come up with unexpected results and want to find out why the outcome was not as predicted. Each experiment usually brings up more unsolved questions than it answers.

This picture of science as a continual growth of new questions is different from that held by many people who think of science as a fixed body of knowledge we need only uncover. This latter concept views scientific research as leaving fewer and fewer questions unanswered as it proceeds. In reality, however, each experiment actually increases the number of questions to be answered.

This open-ended view of science can be very discouraging and very exciting. It can be discouraging because it is sometimes difficult to chart our progress through an ever-expanding universe in which we sometimes seem to take five steps backward for every step forward. On the other hand, it is exciting because we end up asking better and better questions. Perhaps the goal of science is not to find answers to all possible experimental questions but to answer ever more promising and important questions. In following up on your own research, you will find that your main problem is not "How can I get an experimental idea?" but "Which idea is the most important one to work on?"

Focus on a Practical Problem

If you have a practical problem crying for a solution, you may also have the kernel of an experimental idea. So far in this chapter we have been pretending that all research is **basic research**[3]—that is, research done for the sole purpose of increasing the scientific body of knowledge. It is not necessary to justify such research as a solution to any practical problem, although basic research can help solve practical problems. Today's behavior-modification techniques, which provide some of the most powerful procedures for correcting human behavior problems, are based on basic research done in the rat laboratories of yesteryear. Adams (1972) found that many of our military systems were designed using information from basic research done more than 20 years earlier. Thus, basic research does at times prove valuable for solving practical problems. In general, however, such research is done in the name of the advancement of science.

Research designed to solve a specific practical problem is called **applied research**. Perhaps you need to know how humans read handwritten numbers so that you can design a machine to automatically read ZIP codes. Or you may wish to know whether daily quizzes will improve a student's classroom performance on major examinations. Or you may want to know whether transactional analysis is a more effective therapy than psychoanalysis. Many such practical problems need immediate answers that basic researchers might never find. Needing an answer to a practical problem is a perfectly legitimate reason for doing research, and it can be satisfying if your findings have an immediate impact on the world.

Observation is again the key to getting ideas for applied research. Finding a practical problem is simply a matter of carefully observing human behavior and permitting your curiosity free rein. As with the other procedures that we have discussed for getting experimental ideas, you will find that more practical problems need to be solved than you can possibly do experiments to solve. As before, the question becomes "What should I do first?" not "What can I do?"

[3] Basic research is sometimes also called **pure research**, perhaps because one is not supposed to have mixed motives for doing it. Unfortunately, some people who do this kind of research seem to prefer other dictionary definitions of pure—for example, untainted with evil or guilt.

Summary

Although we all have natural curiosity about human behavior, many of us develop irrational fears that block our ideas. Some of us fear that all other researchers are brilliant and that our ideas will not be original. Some people are afraid to propose an experiment requiring complex apparatus, while others are afraid of experiments with simple apparatus. Others fear that their idea is too simple, that their experiment will require complicated statistics, or that their idea is not perfect when it is proposed. Finally, many people do not believe they have good ideas until they translate them into scientific jargonese.

The major key to getting experimental ideas is to learn to **observe** the world about you. You also need to know which ideas are scientifically appropriate. Ideas must be **repeatable, observable,** and **testable** to be experimental ideas. To get ideas you can observe yourself, friends, children, and even pets. Although the best ideas come from direct observation, you can also get ideas by reading other people's research (vicarious observation) and from following up your own research. Finally, some of the best ideas come from having a practical problem that needs to be solved. Research designed to solve a practical problem is called **applied research,** while research aimed primarily at advancing scientific knowledge is called **basic** or **pure research.**

References

Adams, J. A. (1972). Research and the future of engineering psychology. *American Psychologist,* 27, 615-622.

APPENDIX B

How to Be Fair with Science

Adapted from *Doing Psychology Experiments, 5th edition,*
by D. W. Martin. ©2000. Reprinted with permission of Wadsworth, an imprint of the Wadsworth Group, a division of Thomson Learning. Fax 800 730-2215.

Author's note: I have attempted to write Doing Psychology Experiments *in a friendly style. While those who teach psychology have finally admitted that introductory texts need not be written in a pedantic style, we still seem to be unwilling to write about more interesting material in an interesting way. Hopefully this text will be a forerunner of a number of less formal advanced texts.*

In this chapter, we deal with treating science fairly. In some respects, science has fewer defenses than a research participant does. Animals squirm and yell and sometimes die when mistreated, and the animal-rights groups yell, too. Human participants squirm and yell and sometimes sue when mistreated. Science can't even squirm and yell. If you mistreat it long enough, however, your fellow scientists might eventually squirm and yell.

You might wonder how you can be unfair to an inanimate thing like science. In one sense, science *can* be considered animate in that it is a moving, changing, and, we hope, expanding body of knowledge. New research constantly replaces or builds on old findings and theories. Anything you do that retards the expansion of science or causes it to expand in the wrong direction can be considered scientifically unethical.

Science has a few safeguards built into it to ensure that the body of knowledge will continue to expand in a proper direction. For example, before you are allowed to report the outcome of an experiment in the scientific literature, a group of scientists selected for their research accomplishments reviews it. (Think of them as the "Buffy the Vampire Slayers" of the research community.) This review establishes whether the research appears to follow the rules of experimentation discussed in this book. Furthermore, the reviewers attempt to determine whether your contribution is sufficient to warrant using the limited number of pages available in the journals. In this way the reviewers and editors of our journals attempt to screen research so that only competent and relevant findings are added to the body of knowledge. (Note here that I am using the term *relevant* differently than many people use it. I mean relevant to science, not relevant to faddish topics. Sometimes topics that are relevant in the latter sense are the least relevant in the former sense.) Although this reviewing process is not perfect, most psychologists feel that it accomplishes this important screening function rather well.

Although the review system is designed to exclude research that was poorly done or that fails to make a large enough contribution, it is not designed to determine whether an investigator who may be capable of doing good research has lied about his or her research. People who know the rules and say that they have followed them when in fact they have not are cheating science. Such behavior is usually geared toward making personal gains of some sort: "They weren't going to promote me unless I had five publications"; "I had to make our product look good"; or "I had to have a positive result on the experiment to get a passing grade."

Because unethical behavior can be so harmful to our science and because we have so few built-in safeguards to detect it, unethical behavior in science is not tolerated, and a researcher quickly loses the privilege of practicing science when this behavior is detected. I will discuss these clearly unethical and unacceptable behaviors in the first section of this chapter under *dirty tricks.* In the next section I will discuss some behaviors that are frowned upon and that should be guarded against but are not so egregious that a researcher would lose scientific privileges; I will call these *questionable tricks.* Finally, there are times in science when a researcher tells the truth, but not the whole truth, and our science benefits rather than loses from this behavior. This behavior can make our science more efficient and easier to understand and is considered not only acceptable but necessary. I will call these behaviors *neat tricks*.

Dirty Tricks

Fabricating. One form of blatant cheating is to fabricate results. Some "experimenters" know that the easiest way to run an experiment is not to run it at all. They do not have to bother with such mundane matters as buying equipment, signing up participants, or learning to do statistics. All they have to do is learn to write up experiments (fabricators had better read Chapter 13 on How to Report Experimental Results). They had also better learn a different profession, because they will not be psychologists for long.

As a student, you may be tempted to fabricate because an assignment is due and you have not completed it. Don't do it! Late assignments cause lowered grades, but contrived results cause class dismissals and terrible letters of recommendation. Professional scientists are totally intolerant of such behavior.

Back in the early 1900s, a biologist named Kammerer attempted to demonstrate that acquired characteristics could be inherited (Ley, 1955), a concept quite different from Darwinian evolution. He claimed that he had kept generations of fire salamanders on black soil. He reported not only that the salamanders, which are normally black with yellow spots, had shown increasingly smaller spots over generations, but also that this reduction had been passed on by inheritance. A second researcher, who doubted these claims, added the time required to bring forth the number of generations Kammerer had reported and found that the total time was considerably longer than Kammerer had been at work. Other scientists also began to demand explanations until, after seven years, two leading scientists were allowed to examine some specimens. They found injections of India ink, which are not generally considered inherited traits. Kammerer admitted that his results "had plainly been 'improved' post mortem with India ink." (No doubt if he'd pulled a stunt like this in the school yard, some annoying little kid would have yelled, "Liar! Liar! Pants on fire!) He then promptly committed suicide.

The most dastardly deed in science is to add noise to the body of knowledge. If you do bad research, people can and will ignore it, but if you pretend that you have done good research when you have not, you will retard the expansion of the body of knowledge. Others will come along and attempt to build their research on yours, only to discover eventually that something is wrong. They must then waste time fixing the foundation and possibly rebuilding the whole structure. The longer such cheating goes undetected, the greater the eventual waste of science's resources.

If undetected, the cheating of science can also lead to the cheating of society. In the late 1930s, for example, a Russian named T. D. Lysenko also supported the notion that acquired characteristics could be inherited (Lerner, 1968). He was so adamant about this theory that he falsified a great deal of data. Lysenkoites claimed that they had brought about such miraculous results as transforming wheat into rye, barley, oats, and even cornflowers; beets into cabbage; pine into fir; a tree of the hornbeam group into forest walnut (using doctored photographs as evidence); and even the hatching of cuckoos from eggs laid by warblers (Lerner, 1968). Had Lysenko and his cronies lived in later years, they most probably would have gone on to successful careers writing for supermarket tabloids. Just think of the transformations that would have been possible if only aliens could have been added to the mix!

Lysenko's grossly unethical behavior hurt not only science but society as well. He was personally responsible for the dismissal, exile, and execution of a number of Russian geneticists. He convinced Stalin and later Khrushchev that his theories were correct and that they should be applied on a large scale in agricultural programs (Medvedev, 1969). When devastating agricultural failures were eventually attributed in part to Lysenko's methods, he fell, Khrushchev fell (though not only for this reason), and Russian society suffered.

When I wrote this section for the first edition of this book, I had to look back into history to find examples of scientific fraud. Unfortunately, finding examples of fraud is much easier today. In several professional and local newspapers I recently found the following headlines: "Allegations of Plagiarism of Scientific Manuscript Raise Concerns About 'Intellectual Theft,'" "Nobel Honoree Faces Misconduct Charges," and "U.S. Enters Lawsuit Accusing Scientist, Institutions of Fraud."

Are dirty tricks new, or is there simply more interest today in discovering them? Apparently even the old-time scientists were not without guilt. The 19th-century geneticist Gregor Mendel stands accused of having been somewhat less than truthful (Fisher, 1936); apparently he found it easier to run his experiments using only a pencil, rather than skipping between endless rows of pea plants, cross-pollinating with a feather duster as he went. Even Sir Isaac Newton apparently reported correlations to a degree of precision well beyond his ability to measure (Westfall, 1973).

Closer to psychology and more recently, a prominent ESP researcher discovered that his laboratory manager had fudged data. A suspicious research assistant had concealed himself and observed the manager changing the data to support ESP findings. (You would have been able to see the fire on *his* pants for miles!)

The most famous case in psychology concerned Sir Cyril Burt, an eminent British psychologist who had been knighted by King George VI. His research with identical twins was one of the major pillars supporting the argument that IQ is largely inherited. After his death, researchers discovered that the correlations he had reported for identical twins had remained the same to the third decimal place over many years as more twins were added to the study. This seemed a highly unlikely coincidence. Fraud allegations were made in the London *Sunday Times*, and the controversy continues. On one side are researchers who argue that Burt's data should be considered fraudulent and were probably fabricated (McAskie, 1978). Others argue that the case for fraud is weak and that a more probable explanation is simply carelessness (Jensen, 1978). Perhaps Sir Cyril was taking to heart a quote by Samuel Butler: "I do not mind lying, but I hate inaccuracy." It is also unfortunate that this heated debate is fueled by the politics of elitism versus egalitarianism.

How can we guard against the possibility of dishonesty by researchers? No foolproof system will ever be developed, because both the costs and the loss of intellectual freedom would destroy the scientific enterprise (and because, as a society, we seem able to come up with bigger and bigger fools). A number of safeguards are currently in place. The Public Health Service, under which the National Institutes of Health operate, already requires universities to have guidelines for preventing misconduct (Cordes, 1990). The Association of American Universities recommends that institutions establish explicit policies about standards and have administrators to carry out the policies ("AAU Statement," 1983). When the researcher is federally funded, there is even a "whistle-blower" law dating back to the 1800s that allows citizens to file lawsuits against researchers and universities and collect a sizable percentage of any money the government is awarded. Recently, a former technician in a research laboratory filed such a suit against a researcher and his university, alleging that the researcher had falsified results (Cordes, 1990).

One suggestion for minimizing fraud is to establish data archives for the raw data from research (Bryant & Wortman, 1978)—kind of a Fort Knox for data. (The individual measurements of your dependent variable prior to combining them for statistical analysis are your raw data. Unlike meat, raw data do not spoil—they just take up room.) Apparently one problem encountered by investigators attempting to validate Burt's research was his incredibly sloppy data storage. The raw test sheets on the twin studies were among papers stuffed into half a dozen tea chests and later destroyed! The data from research could be stored by individual researchers or possibly sent to a centralized location. In this way, research data would be available to the public, which is often not the case. For instance, a graduate student requesting raw data from 37 published authors received only 24% compliance (Wolins, 1962). Hmmm—it sounds like somebody is hiding something. This may be a case for Mighty Mouse (or Danger Mouse, depending on when you grew up).

The archiving requirement has the following additional benefits: Researchers would probably be more careful in their original data analysis; others could add to the body of knowledge by examining issues not originally addressed; and longitudinal studies in which data are compared over a number of years would be possible.

The costs of a formalized central data storage system include the cost of the bureaucratic structure to administer it, the time and money required to duplicate the data or conform to a standardized format, rent on the small city that would be needed to store the data, and perhaps the cost to science of a recognized loss of trust in the integrity of scientists. Some feel that the costs of this solution are too great when the problem is relatively small. They believe that the checks and balances provided by the replication process are sufficient to detect most fraud and make the problem of dirty tricks an anomaly rather than a common practice.

Regardless of the formal requirements, you should keep data for a minimum of five years (APA, 1994). Investigators often request data from one another, and with computers, storage and retrieval of raw data are easy. If you make a habit of storing raw data, you are helping to protect science, and you are protecting yourself against false accusations as well.

Some data do indicate that scientific fraud is relatively rare. For example, from 1982 to 1988 the National Institutes of Health, which supports the research of 50,000 scientists, had handled reports of only 15 to 20 allegations of wrongdoing. On the other hand, among scientists at a major research university who responded to a survey, one-third said they had suspected a colleague of plagiarism or falsifying data (it was unclear if they all suspected the same colleague), but fewer than half took action to verify or report their suspicions (Hostetler, 1988). At this time it appears that unless scientists take more formal steps to find and eliminate scientific misconduct, Congress will impose such steps through legislation. This possibility disturbs the many scientists who believe that regulation by those not trained or experienced in science could lead to undesirable consequences.

The best and most cost-effective way to prevent fraud in our science is to emphasize to new researchers the importance of ethical conduct. That, of course, is the purpose of the chapter you are reading. It is only through an understanding of the way science works that the real motivation for ethical behavior can be established. Researchers must understand that science is built on trust. Unless we can trust our colleagues' research findings, we might as well discontinue the attempt to build the scientific body of knowledge.

Falsifying Credentials. If you were to walk up to a friend and say, "Would you stand on your head and spin around for me while singing the third verse of the national anthem?" the response would probably be "Why?" However, if you were to walk up to another friend and say, "I'm doing an experiment; would you stand on your head and spin around for me while singing the third verse of the national anthem?" your friend's response is more likely to be "How many times?" (This would depend, of course, on how accommodating your friends are and how often you make such bizarre requests of them.) This difference arises from the fact that our society grants scientists a number of privileges not given to the average citizen. We allow scientists, particularly behavioral scientists, the freedom to experiment because, as a society, we feel that the gains usually outweigh the costs. We also grant scientists a certain amount of prestige and generally respond to them somewhat compliantly.

Scientists not only are allowed to manipulate the lives of those around them in approved ways but also are sometimes supported in this effort by our tax money. However, we are also capable of taking away these privileges if we believe that the gains no longer outweigh the costs. Requiring professional credentials of experimenters is one way we police ourselves. Consequently, to prove to other scientists that you are a qualified investigator, you must present them with your professional record, usually in the form of a vita or resume. A resume is a piece of paper that shows who you are professionally; it lists your educational degrees, your job experience, and your published papers and articles. You use it to get into graduate school, to become professionally certified, or to get a job. Perhaps it should go without saying that this document must be totally accurate. I will say it anyway: *Falsifying credentials* is a blatant dirty trick.

Early in my career, I saw a very talented student attempt to get into graduate school using a falsified vita. He had a fine record and great letters of recommendation from his professors, but he listed several papers and articles on his vita that did not exist. When his professors discovered his deception, the student no longer had either a fine record or letters of recommendation, nor is he an experimental psychologist today. Because the agreement between scientists and society is fragile, this type of dishonesty upsets the delicate balance and cannot be tolerated.

I hope that this discussion of dirty tricks was a waste of your time and that you would not have considered doing these in the first place. Yet I believe that such topics must be mentioned early in an experimenter's training. Doing psychology experiments can be fun, but the real purpose of experimentation is building science. Those who are not willing to follow the rules that make this process orderly do not belong in science.

Questionable Tricks

Those actions that most investigators find unacceptable but that lead to frowning and scolding rather than banishment can be considered questionable tricks. This is the professional equivalent of being rapped on the knuckles with a ruler. These actions can occur during the design of an experiment, during the experiment itself, during data analysis, or in experimental reporting.

Experimental Design. If you design your experiment so that your expectations could cause a desired change in the results and you do not attempt to minimize these effects or even discover them, you are, in a sense, being dishonest. You can also confound an experiment if you claim to have made a particular variable into a control variable when, in fact, this variable systematically changed with your independent variable. In some nonlaboratory experiments such confounding is difficult to control, but in many cases we can legitimately call this situation cheating.

For example, suppose I conducted an experiment in which I attempted to lecture an introductory psychology class at several different rates to determine whether lecture pace had an effect on the students' attentiveness. On some days I tried to speak at a slow pace, on other days at a moderate pace, and on others at a fast pace. We measured attentiveness by recording the level of background noise. Such an experiment would be easy to bias. I could have changed not only my pace but also the degree of liveliness with which I talked about the topic or perhaps the number of interesting examples I used to illustrate my points. Maybe when I lectured quickly I spoke about sex, drugs, and rock and roll, but when I spoke slowly I talked about supply-side economics. These changes in dimensions other than lecture pace could easily confound the independent variable, whether I intended to or not.

One way to minimize the chance of cheating is to design the experiment so that colleagues with little stake in any given outcome rate each lecture in terms of the possible confounding variables. The experimenter can then collect data only from those lectures with equivalent ratings. Moderate cheating in the form of experimenter bias will not necessarily occur in the first design, of course, but the second design will be more convincing because such bias is less likely to occur.

Collecting Data. You can also be dishonest when collecting your data, especially if you must use human judgment to determine what response the participant has made. In the experiment just discussed, for example, the experimenter wants to classify the students' behavior as attentive or inattentive to record the percentage of time spent in attentive versus inattentive listening. Suppose one student sits scribbling with her pencil on a piece of paper. Is she taking notes or doodling? Another student has his eyes closed. Is he concentrating on every word the speaker is saying, or is he just concentrating on the inside of his eyelids? Is a student who is snoring less attentive than one who is merely sleeping? We can classify the behaviors differently depending on our bias. If the experimenter who holds the bias is also doing the classifying, the potential problems are obvious.

To avoid this form of trickery, the experimenter might construct a standard checklist of attentive and inattentive behaviors and have several judges observe the tapes and independently classify the students' behaviors. It is even possible to keep judges blind to the pace the instructor used for the tape being observed. (A blind judge is one who is unaware of which manipulation he or she is viewing. It doesn't mean you have to locate people with visual impairments to help with your research.) Such precautions decrease the possibility of cheating, either intentional or unintentional.

Bias can sometimes occur even in experiments in which measurement of responses seems straightforward. An experiment was carried out in which participants moved a stick to line up a marker with a moving target. After every ten-second interval, the experimenter quickly read a pointer on a voltmeter dial and reset it. (The farther the marker was from the target, the more quickly the pointer moved across the dial.) This task was difficult because the needle seldom fell directly on an index line. The experimenter could easily have made biased judgments about the location of the pointer on the dial. In this experiment, experimenters had to read the dial over 15,000 times, giving rise to the possibility that small inconsistencies in reading the instrument would eventually bias experimental results. (Actually, since the dial had to be read 15,000 times, it was probably either an underpaid graduate student or an undergrad in desperate need of a recommendation for grad school who did the actual recording.) Thus, whenever biased experimenters must use judgment to interpret a response, they should devise procedures to ensure that the judgment will be made accurately.

Data Analysis. You must also avoid trickery in analyzing your data in a biased way. As we will see in Chapter 12, statistical tests are usually computed to determine whether a particular result is likely to be a real effect or whether it is due to chance. These statistical tests can be used only when certain assumptions can be approximated. Using the test when the test assumptions are grossly violated is questionable at best.

For example, the most frequently used statistical tests require that the underlying distribution be approximately normal, a symmetrical bell shape. Although a small violation of this assumption usually does not totally invalidate such a test, some investigators continue to use one of these tests when their distributions in no way resemble a normal distribution. (As a matter of fact, some distributions are so abnormal that it would be considered a step up for them to be called insane!) As an experimenter, it is up to you to know what assumptions your statistical test requires and how likely it is that you will make an error if you fail to meet one or more of these assumptions.

When analyzing your data, you may discover that although most of the participants seem to be showing the predicted experimental effect, several do not show the effect. At this point, you can do nothing about these renegades (unless you are specifically interested in investigating individual differences). Obviously, if you could throw out data from all the participants who failed to show an expected result, you would never do an experiment that failed to support your predictions! For this reason, you must be careful about eliminating participants from an analysis on the basis of their performance on the dependent variable. (Would you disregard abnormal blood pressure readings just because they didn't fit with your image of yourself as a healthy individual?) And you should never eliminate them on the basis of their differential responses to the levels of the independent variable.

You can eliminate participants for failing to meet some overall performance level on the dependent variable only if you determine this performance level before collecting the data, if you can logically defend it, and if you specify the performance level in your experimental report. As an illustration, suppose you were interested in the effects of noise on people's ability to perform a typing task. Before starting the experiment, you might decide to exclude data from all participants who fail to type at least ten words per minute in the absence of noise. Your logic might be that these individuals are such poor typists to begin with that even if noise has a detrimental effect on typing, they would not show the effect. Or you might argue that you are interested in the effect of noise on experienced typists and that a speed of less than ten words per minute indicates that the person is not an experienced typist. However, if you do not have a logical argument for eliminating participants on the basis of a predetermined dependent-variable performance level, you should not do it.

You are much safer in eliminating participants on a basis other than performance on your dependent variable. Again, however, such criteria should be set prior to the experiment and should be specified when you report the results. For example, you might be having participants search through an array of letters to report which letter is printed in red ink. In this case, you might exclude individuals who cannot pass an acuity test or a color-blindness test prior to the experiment.

Reporting Results. Suppose you have analyzed your experiment and are now ready to report the results. Usually you will want to put some of your results into a graph. People have written books on how to lie by *distorting graphs* (Campbell, 1974; Huff, 1954). For example, an experimenter could blow up (that is, enlarge, not demolish) one of the graph's axes to make a tiny effect look like a gigantic effect or possibly distort the scale on one axis so that the function being displayed changes shape. If you are a creative person, you can find all sorts of ways to make even the crummiest, most meaningless, most insignificant results look good. Obviously, such behavior does nothing to advance science and so is considered inappropriate. It's kind of like a doctor adjusting the EKG machine to show normal cardiac rhythm regardless of what's actually happening with the patient

One other form of questionable trickery is *piecemeal reporting* of experimental results. Although research must progress one experiment at a time, you should not report research in this way. Many decades ago the typical journal article in psychology reported the results of a single experiment. In recent times, however, the field has grown so much that there has been a literature explosion. So many people are doing so many experiments that the process of keeping current with experimental advances is nearly impossible. For this reason, few journals will accept a report of a single experiment unless it makes an unusually large contribution by itself. Would you choose a spouse based on only one characteristic? Probably not, unless it was a particularly important characteristic, like having lots and lots of money.

Usually, you should report the results of your experimental research program as an integrated series of experiments. With this procedure, the growth of knowledge becomes much more efficient and orderly, and readers are spared the task of reorienting themselves to the research, rereading introduction and procedure sections with each experiment, and integrating fragmented research into a coherent structure. In today's "publish or perish" world, an investigator can be tempted to do piecemeal reporting to accumulate publications. However, in the end, such behavior does nothing to improve either the investigator's reputation or the body of scientific knowledge.

Neat Tricks

Neat tricks are the little white lies of the scientific community. Although it seems counterintuitive, it is sometimes necessary to "lie" to the reader of a research report to communicate efficiently. Research is usually a sloppy process, yet when you read an experimental report it sounds as if the investigator proceeded in a systematic, orderly fashion at all times. Don't believe it! Rarely does a researcher's mind work in the totally logical manner reflected in the report. Experimenters make many decisions based on hunches or gut-level intuitions. They make false starts based on bad guesses. They do experiments the right way for the wrong reasons or the wrong way for the right reasons. These people are the Inspector Clouseaus of the research world—Clouseau may have been a bumbling idiot, but he always got his man.

Unfortunately, many students become turned off to experimental psychology because they think it is dry and unexciting, when in most cases it is actually an exciting, disorderly, haphazard treasure hunt. You know little about experimentation until you try your first experiment. It's kind of like making love: Reading about it is a poor substitute for doing it.

The most obvious reason for cleaning up an experimental report is to save time and space. Though it might be fun to read about all your colleague's mistakes, you do not have the time and journals do not have the space to allow you the luxury. The experimental report is designed to convey information efficiently, not entertain the reader. (A lot of textbook writers think this, too. They never have any fun!)

Leaving Things Out. One way of cleaning up your experimental report is to leave out experiments and analyses.[1] Suppose you had a horrible, awful, no good, very bad day when you designed the third experiment in a series, you had a bad intuition (for example, suppose you started your morning by putting your milk in a glass and your orange juice on your cereal), or you were temporarily confused. Nobody else is interested in the condition of your life, your viscera, or your head (except maybe your mother—if you're having these problems, call her; she says you don't call often enough as it is). So you blew the experiment. I don't want to read about it. You don't want to write about it. So don't. Science loses nothing, I lose nothing, and you save face. It's a win-win-win situation, which is one win better than just a win-win situation. Be sure, however, that you are not tempted to leave out a perfectly good experiment because the results do not support your favorite hypothesis. Doing this is not a *neat trick*!

Not only is it acceptable to leave out whole experiments if they add nothing to the report, but at times it is also proper to ignore the details of some data analyses. Perhaps there were a number of ways to analyze your data, and you did them all. Although you should probably report that you did the analyses, you need give details of only those that are most representative and convey the most information. (Note that I am not endorsing the practice of conducting a multitude of tests and then picking and choosing only those that yield significant results. In this case, you are distorting the level of significance.) For example, if you conducted a survey with 20 items, you could run 190 comparisons between different pairs of items. You would be extremely hard pressed to find 20 survey items that would be interesting enough for all 190 comparisons to be both meaningful and interesting, so you might as well save yourself (and

[1] Or else put them in a footnote. Nobody reads footnotes.

your potential readers and editors) the time, not to mention the lives of enough trees to cover the state of California that would have to be killed to print out your statistical analyses and interpretations.

Reorganizing. Especially when doing exploratory research, you may find that the outcome of an experiment shows that it should not have been the first experiment in the series. You may find it desirable to back up and do some preliminary experiments. In such cases, you need not tell the reader that "owing to misjudgment, lack of foresight, and overall ineptitude on the part of the dunderheaded experimenter, the following experiments are out of order." You may report them in the most logical order, whether or not this order matches the order in which you did them. Data are data, and you should report them as efficiently as possible, as long as bending the truth does not bend the science.

Reformulating. Finally, it is generally acceptable to *reformulate the theory* underlying an experiment. Occasionally you do an experiment for some reason and later discover a better reason for having done it. Or perhaps you discover that somebody else has done an experiment that casts a different light on one you are conducting. In this case, you have to determine how your contribution to the body of knowledge best fits with the new information. Unfortunately, there will be times when your theory does not fit at all, and you will have to go back to the drawing board. Often, however, you will be able to fit your experiment into the revised theory by changing your emphasis or reinterpreting your results. In reporting your results, you need not burden the reader with obsolete theory. Again, your major ethical consideration should be whether you are adding to the body of knowledge in an efficient manner. Kind of like preparing a paper using a word processor rather than an offset printing press, you want to save yourself and others some time and energy.

In this chapter we have by no means exhausted all the ethical questions that you will face as an experimenter. In some cases, you will find it difficult to decide whether a particular action is fair to science. When a problem comes up, you may wish to discuss it with colleagues, who may be able to raise points and suggest alternatives that you have not considered. In the end, though, the decision is yours. If you apply the principle that ethical actions are those that aid in the efficient growth of the body of knowledge, you will never do dirty tricks and seldom do questionable tricks.

Summary

Because science is a growing body of knowledge, any action that retards the efficient expansion of that body of knowledge is unethical. We can be less than totally truthful with science in a number of ways. We can engage in dirty tricks by *fabricating results* or by *falsifying our credentials.* We can also engage in questionable tricks—by *failing to control confounding variables* in the design of an experiment, for example, or by *misclassifying responses* and *misreading instruments* during collection of our data. During data analysis, *failing to meet test assumptions* and *inappropriately eliminating participants* are also unacceptable, as are distorting graphs and reporting a series of experiments as piecemeal reports. For the sake of efficiency, it is acceptable to write up experimental reports in a form that does not exactly parallel the experiment. For example, we can *leave out experiments and analyses* from a report if they do not add to the report, or we can *reorder experiments* and *reformulate theory* if these actions increase the efficiency of the experimental report.

How to Be Fair with Science

Adapted from *Doing Psychology Experiments, 5th edition*,
by D. W. Martin. ©2000. Reprinted with permission of Wadsworth, an imprint of the Wadsworth Group, a division of Thomson Learning. Fax 800 730-2215.

Author's note: I have attempted to write Doing Psychology Experiments *in a friendly style. While those who teach psychology have finally admitted that introductory texts need not be written in a pedantic style, we still seem to be unwilling to write about more interesting material in an interesting way. Hopefully this text will be a forerunner of a number of less formal advanced texts.*

In this chapter, we deal with treating science fairly. In some respects, science has fewer defenses than a research participant does. Most colleges and universities have committees that make sure animal subjects and human participants aren't mistreated. But they can't ensure that science isn't mistreated. That decision is generally left to your fellow scientists.

You might wonder how you can be unfair to an inanimate thing like science. In one sense, science *can* be considered animate in that it is a moving, changing, and, we hope, expanding body of knowledge. New research constantly replaces or builds on old findings and theories. Anything you do that retards the expansion of science or causes it to expand in the wrong direction can be considered scientifically unethical.

Science has a few safeguards built into it to ensure that the body of knowledge will continue to expand in a proper direction. For example, before you are allowed to report the outcome of an experiment in the scientific literature, a group of scientists selected for their research accomplishments reviews it. This review establishes whether the research appears to follow the rules of experimentation discussed in this book. Furthermore, the reviewers attempt to determine whether your contribution is sufficient to warrant using the limited number of pages available in the journals. In this way the reviewers and editors of our journals attempt to screen research so that only competent and relevant findings are added to the body of knowledge. (Note here that I am using the term *relevant* to mean relevant to science.) Although this reviewing process is not perfect, most psychologists feel that it accomplishes this important screening function rather well.

Although the review system is designed to exclude research that was poorly done or that fails to make a large enough contribution, it is not designed to determine whether an investigator who may be capable of doing good research has lied about his or her research. People who know the rules and say that they have followed them when in fact they have not are cheating science. Such behavior is usually geared toward making personal gains of some sort: "They weren't going to promote me unless I had five publications"; "I had to make our product look good"; or "I had to have a positive result on the experiment to get a passing grade."

Because unethical behavior can be so harmful to our science and because we have so few built-in safeguards to detect it, unethical behavior in science is not tolerated, and a researcher quickly loses the privilege of practicing science when this behavior is detected. I will discuss these clearly unethical and unacceptable behaviors in the first section of this chapter under *dirty tricks*. In the next section I will discuss some behaviors that are frowned upon and that should be guarded against but are not so egregious that a researcher would lose scientific privileges; I will call these *questionable tricks*. Finally, there are times in science when a researcher tells the truth, but not the whole truth, and our science benefits rather than loses from this behavior. This behavior can make our science more efficient and easier to understand and is considered not only acceptable but necessary. I will call these behaviors *neat tricks*.

Dirty Tricks

Fabricating. One form of blatant cheating is to fabricate results. Some "experimenters" know that the easiest way to run an experiment is not to run it at all. They do not have to buy equipment, sign up participants, or learn to do statistics. All they have to do is learn to write up experiments (fabricators had better read Chapter 13 on How to Report Experimental Results). They had also better learn a different profession, because they will not be psychologists for long.

As a student, you may be tempted to fabricate because an assignment is due and you have not completed it. Don't do it! Late assignments cause lowered grades, but contrived results cause class dismissals and terrible letters of recommendation. Professional scientists are totally intolerant of such behavior.

Back in the early 1900s, a biologist named Kammerer attempted to demonstrate that acquired characteristics could be inherited (Ley, 1955), a concept quite different from Darwinian evolution. He claimed that he had kept generations of fire salamanders on black soil. He reported not only that the salamanders, which are normally black with yellow spots, had shown increasingly smaller spots over generations, but also that this reduction had been passed on by inheritance. A second researcher, who doubted these claims, added the time required to bring forth the number of generations Kammerer had reported and found that the total time was considerably longer than Kammerer had been at work. Other scientists also began to demand explanations until, after seven years, two leading scientists were allowed to examine some specimens. They found injections of India ink. Kammerer admitted that his results "had plainly been 'improved' post mortem with India ink." He never recovered his credibility and ended up committing suicide.

As a scientific researcher, the worst thing you can do is to add noise to the body of knowledge. If you do bad research, people can and will ignore it, but if you pretend that you have done good research when you have not, you will retard the expansion of the body of knowledge. Others will come along and attempt to build their research on yours, only to discover eventually that something is wrong. They must then waste time fixing the foundation and possibly rebuilding the whole structure. The longer such cheating goes undetected, the greater the eventual waste of science's resources.

If undetected, the cheating of science can also lead to the cheating of society. In the late 1930s, for example, a Russian named T. D. Lysenko also supported the notion that acquired characteristics could be inherited (Lerner, 1968). He was so adamant about this theory that he falsified a great deal of data. Lysenkoites claimed that they had brought about such miraculous results as transforming wheat into rye, barley, oats, and even cornflowers; beets into cabbage; pine into fir; a tree of the hornbeam group into forest walnut (using doctored photographs as evidence); and even the hatching of cuckoos from eggs laid by warblers (Lerner, 1968).

Lysenko's grossly unethical behavior hurt not only science but society as well. He was personally responsible for the dismissal, exile, and execution of a number of Russian geneticists. He convinced Stalin and later Khrushchev that his theories were correct and that they should be applied on a large scale in agricultural programs (Medvedev, 1969). When devastating agricultural failures were eventually attributed in part to Lysenko's methods, he was discredited and Russian society suffered.

When I wrote this section for the first edition of this book, I had to look back into history to find examples of scientific fraud. Unfortunately, finding examples of fraud is much easier today. In several professional and local newspapers I recently found the following headlines: "Allegations of Plagiarism of Scientific Manuscript Raise Concerns About 'Intellectual Theft,'" "Nobel Honoree Faces Misconduct Charges," and "U.S. Enters Lawsuit Accusing Scientist, Institutions of Fraud."

Are dirty tricks new, or is there simply more interest today in discovering them? Apparently even the old-time scientists were not without guilt. The 19th-century geneticist Gregor Mendel stands accused of having been somewhat less than truthful (Fisher, 1936). Even Sir Isaac Newton apparently reported correlations to a degree of precision well beyond his ability to measure (Westfall, 1973).

Closer to psychology and more recently, a prominent ESP researcher discovered that his laboratory manager had fudged data. A suspicious research assistant had concealed himself and observed the manager changing the data to support ESP findings.

The most famous case in psychology concerned Sir Cyril Burt, an eminent British psychologist who had been knighted by King George VI. His research with identical twins was one of the major pillars supporting the argument that IQ is largely inherited. After his death, researchers discovered that the correlations he had reported for identical twins had remained the same to the third decimal place over many years as more twins were added to the study. This seemed a highly unlikely coincidence. Fraud allegations were made in the London *Sunday Times*, and the controversy continues. On one side are researchers who argue that Burt's data should be considered fraudulent and were probably fabricated (McAskie, 1978). Others argue that the case for fraud is weak and that a more probable explanation is simply carelessness (Jensen, 1978). It is also unfortunate that this heated debate is fueled by the politics of elitism versus egalitarianism.

How can we guard against the possibility of dishonesty by researchers? No foolproof system will ever be developed, because both the costs and the loss of intellectual freedom would destroy the scientific enterprise. A number of safeguards are currently in place. The Public Health Service, under which the National Institutes of Health operate, already requires universities to have guidelines for preventing misconduct (Cordes, 1990). The Association of American Universities recommends that institutions establish explicit policies about standards and have administrators to carry out the policies ("AAU Statement," 1983). When the researcher is federally funded, there is even a "whistle-blower" law dating back to the 1800s that allows citizens to file lawsuits against researchers and universities and collect a sizable percentage of any money the government is awarded. Recently, a former technician in a research laboratory filed such a suit against a researcher and his university, alleging that the researcher had falsified results (Cordes, 1990).

One suggestion for minimizing fraud is to establish data archives for the raw data from research (Bryant & Wortman, 1978). (The individual measurements of your dependent variable prior to combining them for statistical analysis are your raw data.) Apparently one problem encountered by investigators attempting to validate Burt's research was his careless data storage. The raw test sheets on the twin studies were among papers stuffed into half a dozen tea chests and later destroyed! The data from research could be stored by individual researchers or possibly sent to a centralized location. In this way, research data would be available to the public, which is often not the case. For instance, a graduate student requesting raw data from 37 published authors received only 24% compliance (Wolins, 1962).

The archiving requirement has the following additional benefits: Researchers would probably be more careful in their original data analysis; others could add to the body of knowledge by examining issues not originally addressed; and longitudinal studies in which data are compared over a number of years would be possible.

The costs of a formalized central data storage system include the cost of the bureaucratic structure to administer it, the time and money required to duplicate the data or conform to a standardized format, and perhaps the cost to science of a recognized loss of trust in the integrity of scientists. Some feel that the costs of this solution are too great when the problem is relatively small. They believe that the checks and balances provided by the replication process are sufficient to detect most fraud and make the problem of dirty tricks an anomaly rather than a common practice.

Regardless of the formal requirements, you should keep data for a minimum of five years (APA, 1994). Investigators often request data from one another, and with computers, storage and retrieval of raw data are easy. If you make a habit of storing raw data, you are helping to protect science, and you are protecting yourself against false accusations as well.

Some data do indicate that scientific fraud is relatively rare. For example, from 1982 to 1988 the National Institutes of Health, which supports the research of 50,000 scientists, had handled reports of only 15 to 20 allegations of wrongdoing. On the other hand, among scientists at a major research university who responded to a survey, one-third said they had suspected a colleague of plagiarism or falsifying data, but fewer than half took action to verify or report their suspicions (Hostetler, 1988). At this time it appears that unless scientists take more formal steps to find and eliminate scientific misconduct, Congress will impose such steps through legislation. This possibility disturbs the many scientists who believe that regulation by those not trained or experienced in science could lead to undesirable consequences.

The best and most cost-effective way to prevent fraud in our science is to emphasize to new researchers the importance of ethical conduct. That, of course, is the purpose of the chapter you are reading. It is only through an understanding of the way science works that the real motivation for ethical behavior can be established. Researchers must understand that science is built on trust. Unless we can trust our colleagues' research findings, we might as well discontinue the attempt to build the scientific body of knowledge.

Falsifying Credentials. If you were to walk up to a friend and say, "Would you stand on your head for me?" the response would probably be "Why?" However, if you were to walk up to another friend and say, "I'm doing an experiment; would you stand on your head for me?" your friend's response is more likely to be "How long?" This difference arises from the fact that our society grants scientists a number of privileges not given to the average citizen. We allow scientists, particularly behavioral scientists, the freedom to experiment because, as a society, we feel that the gains usually outweigh the costs. We also grant scientists a certain amount of prestige and generally respond to them somewhat compliantly.

Scientists not only are allowed to manipulate the lives of those around them in approved ways but also are sometimes supported in this effort by our tax money. However, we are also capable of taking away these privileges if we believe that the gains no longer outweigh the costs. Requiring professional credentials of experimenters is one way we police ourselves. Consequently, to prove to other scientists that you are a qualified investigator, you must present them with your professional record, usually in the form of a vita or resume. A resume is a piece of paper that shows who you are professionally; it lists your educational degrees, your job experience, and your published papers and articles. You use it to get into graduate school, to become professionally certified, or to get a job. Perhaps it should go without saying that this document must be totally accurate. I will say it anyway: *Falsifying credentials* is a blatant dirty trick.

Early in my career, I saw a very talented student attempt to get into graduate school using a falsified vita. He had a fine record and great letters of recommendation from his professors, but he listed several papers and articles on his vita that did not exist. When his professors discovered his deception, the student no longer had either a fine record or letters of recommendation, nor is he an experimental psychologist today. Because the agreement between scientists and society is fragile, this type of dishonesty upsets the delicate balance and cannot be tolerated.

I hope that this discussion of dirty tricks was a waste of your time and that you would not have considered doing these in the first place.

Yet I believe that such topics must be mentioned early in an experimenter's training. Doing psychology experiments can be fun, but the real purpose of experimentation is building science. Those who are not willing to follow the rules that make this process orderly do not belong in science.

Questionable Tricks

Those actions that most investigators find unacceptable but that lead to reprimands rather than banishment can be considered questionable tricks. These actions can occur during the design of an experiment, during the experiment itself, during data analysis, or in experimental reporting.

Experimental Design. If you design your experiment so that your expectations could cause a desired change in the results and you do not attempt to minimize these effects or even discover them, you are, in a sense, being dishonest. You can also confound an experiment if you claim to have made a particular variable into a control variable when, in fact, this variable systematically changed with your independent variable. In some nonlaboratory experiments such confounding is difficult to control, but in many cases we can legitimately call this situation cheating.

For example, suppose I conducted an experiment in which I attempted to lecture an introductory psychology class at several different rates to determine whether lecture pace had an effect on the students' attentiveness. On some days I tried to speak at a slow pace, on other days at a moderate pace, and on others at a fast pace. We measured attentiveness by recording the level of background noise. Such an experiment would be easy to bias. I could have changed not only my pace but also the degree of liveliness with which I talked about the topic or perhaps the number of interesting examples I used to illustrate my points. These changes in dimensions other than lecture pace could easily confound the independent variable, whether I intended to or not.

One way to minimize the chance of cheating is to design the experiment so that colleagues with little stake in any given outcome rate each lecture in terms of the possible confounding variables. The experimenter can then collect data only from those lectures with equivalent ratings. Moderate cheating in the form of experimenter bias will not necessarily occur in the first design, of course, but the second design will be more convincing because such bias is less likely to occur.

Collecting Data. You can also be dishonest when collecting your data, especially if you must use human judgment to determine what response the participant has made. In the experiment just discussed, for example, the experimenter wants to classify the students' behavior as attentive or inattentive to record the percentage of time spent in attentive versus inattentive listening. Suppose one student sits scribbling with her pencil on a piece of paper. Is she taking notes or doodling? Another student has his eyes closed. Is he concentrating or sleeping? We can classify the behaviors differently depending on our bias. If the experimenter who holds the bias is also doing the classifying, the potential problems are obvious.

To avoid this form of trickery, the experimenter might construct a standard checklist of attentive and inattentive behaviors and have several judges observe the tapes and independently classify the students' behaviors. It is even possible to keep judges blind to the pace the instructor used for the tape being observed. (A blind judge is one who is unaware of which manipulation he or she is viewing.) Such precautions decrease the possibility of cheating, either intentional or unintentional.

Bias can sometimes occur even in experiments in which measurement of responses seems straightforward. An experiment was carried out in which participants moved a stick to line up a marker with a moving target. After every ten-second interval, the experimenter quickly read a pointer on a voltmeter dial and reset it. (The farther the marker was from the target, the more quickly the pointer moved across the dial.) This task was difficult because the needle seldom fell directly on an index line. The experimenter could easily have made biased judgments about the location of the pointer on the dial. In this experiment, experimenters had to read the dial over 15,000 times, giving rise to the possibility that small inconsistencies in reading the instrument would eventually bias experimental results. Thus, whenever biased experimenters must use judgment to interpret a response, they should devise procedures to ensure that the judgment will be made accurately.

Data Analysis. You must also avoid analyzing your data in a biased way. As we will see in Chapter 12, statistical tests are usually computed to determine whether a particular result is likely to be a real effect or whether it is due to chance. These statistical tests can be used only when certain assumptions can be approximated. Using the test when the test assumptions are grossly violated is questionable at best.

For example, the most frequently used statistical tests require that the underlying distribution be approximately normal, a symmetrical bell shape. Although a small violation of this assumption usually does not totally invalidate such a test, some investigators continue to use one of these tests when their distributions in no way resemble a normal distribution. As an experimenter, it is up to you to know what assumptions your statistical test requires and how likely it is that you will make an error if you fail to meet one or more of these assumptions.

When analyzing your data, you may discover that although most of the participants seem to be showing the predicted experimental effect, several do not show the effect. At this point, you can do nothing about these renegades (unless you are specifically interested in investigating individual differences). Obviously, if you could throw out data from all the participants who failed to show an expected result, you would never do an experiment that failed to support your predictions! For this reason, you must be careful about eliminating participants from an analysis on the basis of their performance on the dependent variable. And you should never eliminate them on the basis of their differential responses to the levels of the independent variable.

You can eliminate participants for failing to meet some overall performance level on the dependent variable only if you determine this performance level before collecting the data, if you can logically defend it, and if you specify the performance level in your experimental report. As an illustration, suppose you were interested in the effects of noise on people's ability to perform a typing task. Before starting the experiment, you might decide to exclude data from all participants who fail to type at least ten words per minute in the absence of noise. Your logic might be that these individuals are such poor typists to begin with that even if noise has a detrimental effect on typing, they would not show the effect. Or you might argue that you are interested in the effect of noise on experienced typists and that a speed of less than ten words per minute indicates that the person is not an experienced typist. However, if you do not have a logical argument for eliminating participants on the basis of a predetermined dependent-variable performance level, you should not do it.

You are much safer in eliminating participants on a basis other than performance on your dependent variable. Again, however, such criteria should be set prior to the experiment and should be specified when you report the results. For example, you might be having participants search through an array of letters to report which letter is printed in red ink. In this case, you might exclude individuals who cannot pass an acuity test or a color-blindness test prior to the experiment.

Reporting Results. Suppose you have analyzed your experiment and are now ready to report the results. Usually you will want to put some of your results into a graph. People have written books on how to lie by *distorting graphs* (Campbell, 1974; Huff, 1954). For example, an experimenter could blow up one of the graph's axes to make a tiny effect look like a gigantic effect or possibly distort the scale on one axis so that the function being displayed changes shape. If you are a creative person, you can find all sorts of ways to make even crummy results look good. Obviously, such behavior does nothing to advance science and so is considered inappropriate.

One other form of questionable trickery is *piecemeal reporting* of experimental results. Although research must progress one experiment at a time, you should not report research in this way. Many decades ago the typical journal article in psychology reported the results of a single experiment. In recent times, however, the field has grown so much that there has been a literature explosion. So many people are doing so many experiments that the process of keeping current with experimental advances is nearly impossible. For this reason, few journals will accept a report of a single experiment unless it makes an unusually large contribution by itself.

Usually, you should report the results of your experimental research program as an integrated series of experiments. With this procedure, the growth of knowledge becomes much more efficient and orderly, and readers are spared the task of reorienting themselves to the research, rereading introduction and procedure sections with each experiment, and integrating fragmented research into a coherent structure. In today's "publish or perish" world, an investigator can be tempted to do piecemeal reporting to accumulate publications. However, in the end, such behavior does nothing to improve either the investigator's reputation or the body of scientific knowledge.

Neat Tricks

Although it seems counterintuitive, it is sometimes necessary to "lie" to the reader of a research report to communicate efficiently. Research is usually a sloppy process, yet when you read an experimental report it sounds as if the investigator proceeded in a systematic, orderly fashion at all times. Don't believe it! Rarely does a researcher's mind work in the totally logical manner reflected in the report. Experimenters make many decisions based on hunches or gut-level intuitions. They make false starts based on bad guesses. They do experiments the right way for the wrong reasons or the wrong way for the right reasons.

Unfortunately, many students become turned off to experimental psychology because they think it is dry and unexciting, when in most cases it is actually a treasure hunt. You know little about experimentation until you try your first experiment.

The most obvious reason for cleaning up an experimental report is to save time and space. Though it might be fun to read about all your colleague's mistakes, you do not have the time and journals do not have the space to allow you the luxury. The experimental report is designed to convey information efficiently, not entertain the reader.

Leaving Things Out. One way of cleaning up your experimental report is to leave out experiments and analyses. Suppose you had a bad day when you designed the third experiment in a series, you had a bad intuition, or you were temporarily confused. So you blew the experiment. I don't want to read about it. You don't want to write about it. So don't. Science loses nothing, I lose nothing, and you save face. Be sure, however, that you are not tempted to leave out a perfectly good experiment because the results do not support your favorite hypothesis. Doing this is not a *neat trick*!

Not only is it acceptable to leave out whole experiments if they add nothing to the report, but at times it is also proper to ignore the details of some data analyses. Perhaps there were a number of ways to analyze your data, and you did them all. Although you should probably report that you did the analyses, you need give details of only those that are most representative and convey the most information. (Note that I am not endorsing the practice of conducting a multitude of tests and then picking and choosing only those that yield significant results. In this case, you are distorting the level of significance.) For example, if you conducted a survey with 20 items, you could run 190 comparisons between different pairs of items. You would be extremely hard pressed to find 20 survey items that would be interesting enough for all 190 comparisons to be both meaningful and interesting, so you might as well save yourself (and your potential readers and editors) the time.

Reorganizing. Especially when doing exploratory research, you may find that the outcome of an experiment shows that it should not have been the first experiment in the series. You may find it desirable to back up and do some preliminary experiments. In such cases, you need not tell the reader that "owing to misjudgment and lack of foresight of the experimenter, the following experiments are out of order." You may report them in the most logical order, whether or not this order matches the order in which you did them. Data are data, and you should report them as efficiently as possible, as long as bending the truth does not bend the science.

Reformulating. Finally, it is generally acceptable to *reformulate the theory* underlying an experiment. Occasionally you do an experiment for some reason and later discover a better reason for having done it. Or perhaps you discover that somebody else has done an experiment that casts a different light on one you are conducting. In this case, you have to determine how your contribution to the body of knowledge best fits with the new information. Unfortunately, there will be times when your theory does not fit at all, and you will have to go back to the drawing board. Often, however, you will be able to fit your experiment into the revised theory by changing your emphasis or reinterpreting your results. In reporting your results, you need not burden the reader with obsolete theory. Again, your major ethical consideration should be whether you are adding to the body of knowledge in an efficient manner.

In this chapter we have by no means exhausted all the ethical questions that you will face as an experimenter. In some cases, you will find it difficult to decide whether a particular action is fair to science. When a problem comes up, you may wish to discuss it with colleagues, who may be able to raise points and suggest alternatives that you have not considered. In the end, though, the decision is yours. If you apply the principle that ethical actions are those that aid in the efficient growth of the body of knowledge, you will never do dirty tricks and seldom do questionable tricks.

Summary

Because science is a growing body of knowledge, any action that retards the efficient expansion of that body of knowledge is unethical. We can be less than totally truthful with science in a number of ways. We can engage in dirty tricks by *fabricating results* or by *falsifying our credentials.* We can also engage in questionable tricks—by *failing to control confounding variables* in the design of an experiment, for example, or by *misclassifying responses* and *misreading instruments* during collection of our data. During data analysis, *failing to meet test assumptions* and *inappropriately eliminating participants* are also unacceptable, as are distorting graphs and reporting a series of experiments as piecemeal reports. For the sake of efficiency, it is acceptable to write up experimental reports in a form that does not exactly parallel the experiment. For example, we can *leave out experiments and analyses* from a report if they do not add to the report, or we can *reorder experiments* and *reformulate theory* if these actions increase the efficiency of the experimental report.

APPENDIX C

How to Get an Experimental Idea

1. Martin listed nine reasons why students have trouble coming up with experimental ideas. What are they? (9 points)
   - Geniephobia—fear of geniuses
   - Imitatophobia—fear of imitating
   - Paraphernaliophobia—fear of apparatus
   - Manuphobia—fear of doing things by hand
   - Parsimoniophobia—fear of simplicity
   - Calculatophobia—fear of statistics
   - Imperfectaphobia—fear of being imperfect
   - Pseudomonophonoscientiaphobia—fear of not sounding scientific
   - Ergophobia—fear of work

2. According to Martin, what are the two factors that most strongly influence the ability to come up with a good experimental idea? (2)
   - Observation
   - Curiosity

3. List the five people or groups Martin suggests that you observe directly as a means of getting experimental ideas. (2½)
   - Yourself
   - Friends
   - Children
   - Pets
   - Others/public

4. What is vicarious observation? (2)
   - Getting experimental ideas by reading other people's research

5. What is basic research? (1)
   - Research done for the sole purpose of increasing the scientific body of knowledge

6. According to Martin, what are three ways you might expand on your own research? (1½)
   - Answer questions raised by your results
   - Set a variable to a different level in a new experiment
   - Explore unexpected results

7. What are introspectionists? (1)
   - Researchers who concentrated on looking at their own mental processes

How to Get an Experimental Idea

1. According to David Martin, what is the main reason that students believe they cannot come up with experimental ideas? (1 point)
   a. They are afraid there is something wrong with their ideas.*
   b. They do not have enough knowledge to come up with good ideas.
   c. They lack the scientific knowledge needed to come up with good ideas.
   d. They lack the necessary life experience to come up with ideas..

2. A student reads an article about social learning among primates—specifically, the article discusses how baboons can learn how to fish termites from a termite hill using a stick by watching another baboon do it successfully. The student wonders if pigeons can also learn food gathering techniques through social learning. This an example of getting an experimental idea through: (1)
   a. direct observation.
   b. vicarious observation.*
   c. expanding on one's research.
   d. focusing on a practical problem.

3. What does Martin encourage as the main source for getting new experimental ideas? (1)
   a. Studying the research of others for inspiration
   b. Focusing on a practical problem that needs to be solved
   c. Using your prior research as a starting point for future studies
   d. Observing the behavior of yourself and others*

4. Why does Martin advocate using your own research as a means for getting ideas? (1)
   a. You will already be familiar with the literature, so it will be easier than starting over with a new idea.
   b. It's better for your professional development to know a lot about a fairly specific area than a little about a lot of topics.
   c. You will probably have more questions than answers after conducting a research study.*
   d. Once you are an expert in a specific field, you are less likely to be challenged by your colleagues.

5. Which of these is an example of someone getting an idea through direct observation? (1)
   a. A student reads a journal article and decides to pursue one of the questions the researchers raised as a result of their experiments.
   b. A student wonders if a certain memory technique that relies on visualization will work more effectively with people who are visual learners, as opposed to people who learn better through auditory or kinesthetic techniques.

c. A student does a study and finds that using cartoons in a presentation helps maintain audience attention. The student decides to explore the effect of differing numbers of cartoons in presentations of a certain length.
d. A student who works in a day care center wonders if girls or boys will show a greater increase in aggressive behavior after watching cartoon violence.*

6. Using the three characteristics of an experimental idea, which of these describes the best experimental idea? (1)
   a. Are people who wear glasses perceived as smarter?*
   b. Do animals dream in color?
   c. Is assisted suicide for terminally ill cancer patients wrong?
   d. Can some people communicate psychically with animals?

7. What is basic research? (1)
   a. Research that solves a problem
   b. Research that is original and builds on little if any previous research
   c. Research that is done solely to increase the body of knowledge*
   d. Research that that uses little or no apparatus

8. Which of these is an example of someone focusing on a practical problem to come up with an experimental idea? (1)
   a. A student who is a part-time cashier wonders if men or women are more likely to pay for their purchases with cash.
   b. A student wonders if positive reinforcement works equally well for different kinds of animals.
   c. A student found in a previous study that subjects recall more unrelated items from a list using a particular memory aid and now wonders how subjects can be encouraged to put this memory aid to use for school work.
   d. A student who works in a day-care center wonders how to best organize the play equipment so all the children get sufficient exercise.*

9. A woman observes that when she talks on the phone, she frequently engages in another activity, such as paying bills. Her boyfriend, however, paces the floor while talking on the phone. The woman wonders if there is a difference between men and women in their "auxiliary" behaviors on the telephone.
   This an example of getting an experimental idea through: (1)
   a. direct observation.*
   b. vicarious observation.
   c. expanding on one's research.
   d. focusing on a practical problem.

How to Be Fair with Science

1. How does Martin define "dirty tricks"? (2)
   - Unethical and unacceptable behaviors

2. Martin describes two dirty tricks. What are they? (2)
   - Fabricating
   - Falsifying credentials

3. How does Martin define "questionable tricks"? (2)
   - Behaviors that are frowned on and that should be guarded against, but are not so egregious that a researcher would lose scientific privileges

4. Martin lists four stages during which questionable tricks may occur. What are they? (4)
   - Experimental design
   - Collecting data
   - Data analysis
   - Reporting results

5. How does Martin define "neat tricks"? (2)
   - Behaviors that make science more efficient and easier to understand

6. Martin describes three neat tricks. What are they? (3)
   - Leaving things out
   - Reorganizing
   - Reformulating

7. What is the purpose of a review of a scientific article prior to publication? (2)
   - Establish whether the research appears to follow the rules of experimentation
   - Determine whether contribution is sufficient to warrant using the limited number of pages in the journal

8. What does Martin say are the four benefits of establishing a data archive? (2)
   - Minimizes fraud
   - Researchers would be more careful in their original data analysis
   - Other people could add to the body of knowledge by examining issues not originally addressed
   - Longitudinal studies in which data are compared over time would be possible

9. According to Martin, what is the best and most cost-effective way to prevent fraud in science? (1)
   - Emphasize to new researchers the importance of ethical conduct

How to Be Fair with Science

1. What is the purpose of a review of a scientific article prior to publication? (I point)
   a. Determine if the authors have the necessary credentials to be considered credible
   b. Determine if the writing style is appropriate for the journal
   c. Determine whether an investigator who is capable of doing good research has lied about his/her findings
   d. Establish whether the research appears to follow the rules of experimentation*

2. What is the likely penalty to a student or researcher who falsifies his/her credentials? (1)
   a. A lawsuit
   b. Criminal charges
   c. Dismissal from the field*
   d. A full audit of research by a national organization such as the National Institutes of Health or the Association of American Universities

3. According to Martin, what is the danger of scientists failing to take formal steps to find and eliminate scientific misconduct? (1)
   a. Scientific misconduct will become increasingly prevalent.
   b. Scientists will lose public funding for their research.
   c. Congress may impose such steps through legislation.*
   d. The general public will lose confidence in the results of research.

4. Suppose you are studying how morning television affects children's behavior. Which of the following situations would Martin consider to be cheating? (1)
   a. You discover that the children who watched Mighty Morphin Power Rangers had Chocolate Frosted Sugar Bombs and doughnuts for breakfast, while children who watched Barney ate a breakfast of eggs and ham. You decide that the breakfast probably didn't make any difference and decide not to consider that as a factor in your study.*
   b. As you are writing up the results of your study, you find that an article in a new issue of a journal you read poses a theory that better explains your data than the theory you had originally operated under. You decide to replace your old theory with this new one and write up your results in light of the new theory.
   c. You use a questionnaire that has 58 items for each child, many of which you find to have no relevance to the data you collected. You decide against reporting the results of all 58 items and any possible analysis that could be done using that data.
   d. You conduct this study as one in a series of studies. You find that when you are done, it seems to make more sense to present the results of the studies in an order that is different from the chronological order in which the studies

were conducted. You decide to present the studies in an order other than the one in which they were conducted.

5. A researcher is comparing two methods of teaching reading: phonics and whole language. In which of the following situations would it be okay for the researcher to remove subjects from the study? (1)
   a. Some children in both groups failed to show any improvement in reading. Since the results of these individuals fail to either confirm or deny the hypothesis, the researcher can safely disregard data from these subjects.
   b. Some of the subjects in the phonics group showed an unexpectedly large increase in reading skills. If the researcher feels that these children may have received supplemental instruction that enhanced their learning, the data from these subjects may be disregarded.
   c. Some of the subjects read proficiently prior to the study. Since the researcher decided prior to collecting data that proficient readers would not show much improvement and so would not contribute to the understanding of teaching methods for reading, it is okay to disregard data for these subjects.*
   d. It is never okay to remove subjects from a study after collecting data.

6. According to Martin, what is the best and most cost-effective way to prevent fraud in science? (1)
   a. Emphasize to new researchers the importance of ethical conduct*
   b. Require internal review boards to perform random audits on research conducted at their facility
   c. Encourage funding agencies to perform random audits on research they fund
   d. Require researchers to sign affidavits regarding their adherence to ethical standards

7. According to Martin, which of the following is ***not*** a questionable behavior? (1)
   a. You hire a student to record the amount of time a pair of young rats exhibit "play behavior." You tell the student what the hypothesis of the study is and which animals are in which experimental group, but you leave it up to the student to determine exactly what play behavior is and how it should be recorded.
   b. You hire a student to record the amount of time a pair of young rats exhibit "play behavior." You provide the student with training on what constitutes "play behavior" in young rats. You also tell the student what the hypothesis of the experiment is, and which animals are in which experimental condition.
   c. You realize that the majority of your data on play behavior shows that the rats spent either a large amount of time playing or very little time playing. The statistical test you intended to use requires that data be in the shape of a bell. You decide to use that test.
   d. You have a graduate student record the amount of time a pair of young rats exhibit "play behavior." You provide the graduate student with training on what constitutes "play behavior" in young rats. You don't tell the student what

the hypothesis of the experiment is, or which animals are in which experimental condition.*

For each of the following, indicate if it is something Martin referred to as a "dirty trick," a "questionable trick," or a "neat trick." (1 point each)

__B__ 8. A researcher wants to test a new method of teaching physical education skills to elementary students. He has the phys ed teacher use the old method in the classes that meet before lunch and the new method for the classes that meet after lunch. __B__ 9. An experimenter spends extra time with subjects in a group he expects to perform better, and less time with subjects he expects to perform poorly. __C__ 10. After running a study on learning, a scientist decides that the study is more similar to studies that use situated learning theory instead of the cognitive learning theory that he had in mind when he ran the study. He decides to write up the study using situated learning theory as its basis. __A__ 11. A scientist beefs up his resume by including a few fictitious papers. __B__ 12. A scientist uses a statistical test that requires a bell curve distribution, even though he knows his data is not distributed in a bell curve. __C__ 13. A scientist runs some data analyses that reveal no new information, so she decides to leave them out of her report. __C__ 14. After running several related experiments, the research decided that it makes more sense to report them in an order other than the order in which they were run. __A__ 15. A researcher fails to receive funding for her research, but is still under pressure from her employer to publish articles. She runs a few subjects, creates data for the remaining subjects, and then publishes the results of her "research."	a. dirty trick b. questionable trick c. neat trick

__B__ 16. A researcher is investigating the effect of sugary snacks on the behavior of preschool children. He observes tapes of the children and records their behaviors. He is aware of which children have had sugary snacks and which have had healthier snacks.	
__C__ 17. A researcher gets unexpected results, but is able to explain them using a different theory than the one he had in mind when he designed the study. He decides to report the results in light of this other theory.	
__B__ 18. A researcher wants to study a new teaching method. He has all the history teachers in the high school use the old method and all the science teachers use the new method.	a. dirty trick b. questionable trick
__C__ 19. After running a few related studies, a researcher decides that he should have run another study to help provide the groundwork for the other studies. After running this final study, he writes up the report presenting the last study first.	c. neat trick
__B__ 20. When a researcher runs an experiment, she notices that the data from a few of the subjects are substantially different from that of the vast majority of subjects. She assumes these data are not reflective of the majority of the subjects and eliminates them from her analyses.	