

## **INFORMATION TO USERS**

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

# **UMI**

A Bell & Howell Information Company  
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA  
313/761-4700 800/521-0600



**RETROSPECTIVE PRETESTS: CONCEPTUAL AND METHODOLOGICAL ISSUES**

by

**Judith Lynn Babcock**

---

**Copyright © Judith Lynn Babcock 1997**

**A Dissertation Submitted to the Faculty of the**

**DEPARTMENT OF PSYCHOLOGY**

**In partial Fulfillment of the Requirements  
For the Degree of**

**DOCTOR OF PHILOSOPHY**

**In the Graduate College**

**THE UNIVERSITY OF ARIZONA**

**1997**

**UMI Number: 9806822**

**Copyright 1997 by  
Babcock, Judith Lynn**

**All rights reserved.**

---

**UMI Microform 9806822  
Copyright 1997, by UMI Company. All rights reserved.**

**This microform edition is protected against unauthorized  
copying under Title 17, United States Code.**

---

**UMI**  
**300 North Zeeb Road**  
**Ann Arbor, MI 48103**

THE UNIVERSITY OF ARIZONA ©  
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Judith Lynn Babcock entitled Retrospective Pretests: Conceptual and Methodological Issues

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

<u>Lee Sechrest</u> Lee Sechrest, PhD	<u>6/9/97</u> Date
<u>Dick Bootzin</u> Dick Bootzin, PhD	<u>6/9/97</u> Date
<u>Scott Leischo</u> Scott Leischo, PhD	<u>6/10/97</u> Date
<u>A. J. Figueredo</u> A. J. Figueredo, PhD	<u>6/9/97</u> Date
<u>Julie Erickson</u> Julie Erickson, PhD	<u>6/9/97</u> Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

<u>Lee Sechrest</u> Dissertation Director, Lee Sechrest, PhD	<u>6/9/97</u> Date
---	-----------------------

**STATEMENT BY AUTHOR**

**This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.**

**Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.**

**SIGNED:** 

## TABLE OF CONTENTS

LIST OF ILLUSTRATIONS .....	6
LIST OF TABLES .....	7
ABSTRACT.....	8
<b>CHAPTER 1: INTRODUCTION .....</b>	<b>9</b>
<b>Aim .....</b>	<b>9</b>
<b>Background .....</b>	<b>9</b>
<b>The Retrospective Pretest .....</b>	<b>10</b>
<b>Retrospective Pretests and the Response Shift Phenomenon .....</b>	<b>12</b>
<b>Response-Shift Bias and the Validity of Retrospective Pretests .....</b>	<b>17</b>
<b>Sources of Bias in Retrospective Pretests .....</b>	<b>21</b>
<b>Response-Style Effects .....</b>	<b>21</b>
<i>Demand Characteristics .....</i>	<i>21</i>
<i>Motivation .....</i>	<i>27</i>
<b>Awareness of an Objective Measure .....</b>	<b>28</b>
<b>Hindsight Bias .....</b>	<b>31</b>
<b>Order Effects .....</b>	<b>31</b>
<b>Memory Distortion in Retrospective Pretests .....</b>	<b>33</b>
<i>Autobiographical Recall .....</i>	<i>37</i>
<i>Complex Episodes .....</i>	<i>39</i>
<i>Personal Attributes .....</i>	<i>39</i>
<i>Bias Associated with Variable Types.....</i>	<i>41</i>
<i>Attitudinal Variables .....</i>	<i>41</i>
<i>Cognitive Variables.....</i>	<i>43</i>
<i>Variable Complexity .....</i>	<i>45</i>
<b>Summary .....</b>	<b>47</b>
<b>Present Study .....</b>	<b>48</b>
<b>Research Questions .....</b>	<b>49</b>
<b>CHAPTER 2: PHASE 1: IDENTIFICATION AND CROSS-VALIDATION OF THE VARIABLE TYPES .....</b>	<b>52</b>
<b>New Student Survey .....</b>	<b>52</b>
<b>Method .....</b>	<b>53</b>
<b>Data Analysis .....</b>	<b>54</b>
<b>Results and Discussion .....</b>	<b>54</b>
<i>Factor Confirmation .....</i>	<i>54</i>
<i>Cross-Validation .....</i>	<i>56</i>
<b>Beck Depression Inventory .....</b>	<b>56</b>
<b>Method.....</b>	<b>57</b>

**TABLE OF CONTENTS - *Continued***

<b>Data Analysis</b> .....	57
<b>Results and Discussion</b> .....	58
<b>Cross-Validation</b> .....	58
<b>CHAPTER 3: PHASE 2: RETROSPECTIVE PRETEST STUDY</b> .....	60
<b>Study Design</b> .....	60
<b>Method</b> .....	60
<b>Participants</b> .....	60
<b>Procedure</b> .....	60
<b>Data Analysis</b> .....	61
<b>Results and Discussion</b> .....	63
<b>Difference Scores</b> .....	63
<b>Effect of Time</b> .....	66
<b>Structural Equations Models</b> .....	67
<b>Individual Items</b> .....	71
<b>General Discussion</b> .....	74
<b>CHAPTER 4: CONCLUSIONS</b> .....	76
<b>APPENDIX A: NEW STUDENT SURVEY</b> .....	80
<b>APPENDIX B: BECK DEPRESSION INVENTORY</b> .....	83
<b>APPENDIX C: RETRO STUDENT SURVEY</b> .....	86
<b>APPENDIX D: POST STUDENT SURVEY</b> .....	90
<b>REFERENCES</b> .....	93

**LIST OF ILLUSTRATIONS**

<b>FIGURE 1. Analytical model used to test for memory distortion in retrospective pretests.....</b>	<b>54</b>
<b>FIGURE 2. Analytical model of the Optimism variable type .....</b>	<b>69</b>
<b>FIGURE 3. Analytical model of the Ability variable type .....</b>	<b>70</b>
<b>FIGURE 4. Analytical model of the Mood variable type .....</b>	<b>71</b>

## LIST OF TABLES

TABLE 1. Memory distortion in retrospective data across time.....	47
TABLE 2. Internal consistency estimates for the Optimism factor.....	55
TABLE 3. Internal consistency estimates for the Ability factor.....	55
TABLE 4. Internal consistency estimates for the Mood factor.....	56
TABLE 5. Cross-validation of the Optimism, Ability, and Health factors.....	56
TABLE 6. Internal consistency estimates for the 9-item BDI scale.....	58
TABLE 7. Cross-validation of the BDI 9-item scale.....	59
TABLE 8. Data collection plan.....	60
TABLE 9. Coefficient alpha for the four variable types over time.....	64
TABLE 10. Post-minus-pre difference scores across time.....	65
TABLE 11. Post-minus-retro difference scores across time.....	65
TABLE 12. Retro-minus-pre difference scores across time.....	66
TABLE 13. Statistical and practical fit indices for the multisample analysis.....	68
TABLE 14. Testing for hindsight bias.....	73
TABLE 15. Accuracy of retrospective pretest reports for specific items.....	73

## ABSTRACT

Retrospective pretests provide a potentially useful elaboration on research methodology if they can be shown to be dependable under specific sets of conditions. Previous studies have examined response-shift bias and response-style effects, but less attention has been given to memory distortion associated with the retrospective recall of diverse types of variables. Identifying psychometric characteristics of these measures may help to clarify the picture emerging from retrospective accounts.

The present study applied a methodology developed to measure the systematic error (i.e., memory distortion) that may be associated with variables involving a range of recall tasks. The study examined which types of variables account for the least measurement error in retrospective pretests administered at three time points. The types of variables examined in this study include students' self-ratings of academic abilities, self-reported attitudes and opinions about college, mood states, and perceptions of general health.

The results of this study indicate that there was no main effect of time on any of the pairs of difference scores, and a moderate level of memory distortion was detected in the three variable types examined. The methodology applied provides an effective approach to understanding the effect of memory distortion on retrospective pretest variables. The author recommends that future applications of this methodology be applied to heterogeneous populations, investigate a range of complex variables, and include an examination of individual subject differences.

## CHAPTER 1: INTRODUCTION

### Aim

Retrospective pretests provide a potentially useful elaboration on research methodology if they can be shown to be dependable under specific sets of conditions. Previous studies have examined response-shift bias and response-style effects, but less attention has been given to memory distortion that is associated with recall of diverse types of variables used in retrospective pretests (e.g., opinions, attitudes, cognitions). Identifying psychometric characteristics of these measures may help to clarify the picture emerging from retrospective accounts. To assist in this identification process, a more systematic method of research is needed to investigate the effect that instrumentation may have on retrospective responses. The aim of the present study is to identify sources of systematic error (i.e., memory distortion) that may occur in the recall of diverse variables types used in retrospective pretests. Specifically, the study is designed to determine which types of variables account for the least measurement error in retrospective pretests administered at 4, 8, and 12 week time points. The types of variables examined in this study include students' self-ratings of academic abilities, self-reported attitudes and opinions about college, mood states, and perceptions of general health.

### Background

Researchers have an array of research design options available for assessing change, and each option possesses inherent strengths and weaknesses. Methodologists suggest that it is not uncommon for the reliability and superiority of one design option to vanish in a given population or under inescapable circumstances that introduce a variety of

practical problems. For many options, parameters of application and issues concerning reliability and validity are fairly straightforward. The case for the retrospective pretest option is not so easily determined.

### The Retrospective Pretest

Campbell and Stanley (1963) first recommended the use of retrospective pretests as a means for supplementing data to rule out plausible rival hypotheses. The retrospective pretest method requires subjects to complete a posttest following an intervention or training. Subjects are then asked to give an additional report concerning their level of functioning on the construct of interest prior to the training or intervention. For example, upon completion of a communication skills training program, participants may be asked to complete a self-report questionnaire that requires them to rate their communication skills. When the posttest is completed, participants are then asked to complete another questionnaire, a retrospective pretest, but this time they are to indicate what their communication skills were like prior to the workshop. The retrospective pretest thus provides an opportunity to obtain ratings that are in closer proximity to the posttest. A respondent may or may not complete a pretest, however, but if one is completed, then the retrospective pretest should not be an attempt by respondents to remember what was initially reported on the pretest.

The retrospective pretest initially gained the attention of the research community when Campbell and Stanley referred to an analysis conducted by Deutsch and Collins (1951) in which retrospective pretests were used to examine differences in the attitudes of housing project occupants in integrated versus segregated units. At the time of the study,

only posttest measures were available, and the investigators feared the differences they found might be regarded as reflecting selection biases in initial attitudes. Using the retrospective pretest, however, they found no differences between the two types of housing groups in remembered prior attitudes. Although a pretest entrance interview and random assignment of tenants to treatments would have been a much stronger research design, Campbell and Stanley (1963) concluded that the retrospective pretest was a “precious contribution to an experimentally oriented science” (p.66).

During the sixties and seventies, the research design most frequently used to evaluate change was Campbell and Stanley’s (1963, Design 4) two-group pretest-posttest design. Although that design provides an opportunity to measure change attributed to training or treatment, there are also caveats surrounding its use. One caveat noted by Cronbach and Furby (1970) indicated that in order for pretest and posttest scores to be comparable, a common metric must exist, and that requirement was not always met in applications of the pretest-posttest design. Another cautionary note by Howard, Ralph, Gulanick, Maxwell, Nance and Gerber (1979) warned that using self-report instruments may result in a “response-shift” or a change in the internal standard from pretest to posttest, yielded incompatible pre-posttest scores.

Failure to find differences between self-report pre- and posttest measurements became a concern for many researchers. For instance, when researchers tried to measure effects of psychological treatments using self-report measures, they failed to find convincing support for the continuation of treatment even though therapists believed that the treatment had been beneficial (Bergen, 1971). The discrepancy between research

findings and both client and therapist perceptions resulted in many questions about the adequacy of research methods and instruments to assess psychological benefits (Howard, Ralph, et al. 1979). Other researchers such as Pohl (1982) noted that many investigators often shy away from the self-report technique because of its assumed invalidity. In response to these warnings, numerous studies began employing the retrospective pretest in an attempt to correct for diverse sources of error. It is also likely that the use of retrospective pretests reflects their low cost, the fact that they do not have to be planned in advance, and other methodological compromises.

#### Retrospective Pretests and the Response Shift Phenomenon

Soon after the retrospective pretest was employed as an alternative approach to data collection, researchers attempted to understand the discrepancies reported between pretests and retrospective pretests. There are several possible explanations for the discrepancies that must be considered.

First, when observed change is attributed to true change, there should be no discrepancy between pretest and retrospective pretest reports. For example, people enrolled in a weight loss program may report their weight when the program begins (i.e., at pretest). Three months later, one might ask them to report both their current weight (i.e., posttest) and their weight when they began the program (i.e., retrospective pretest). If they are able to provide an accurate report, then the true weight change can be determined by either using the post-minus-prescores or the post-minus-retrospective scores. We know that measuring "true change" is not always possible, however. In some treatment or training programs, for instance, people may wish to create the impression of

having changed or not changed, thereby producing a discrepancy between pre- and retrospective reports. For example, an employee enrolled in a required leadership workshop who wants to make a positive impression on his supervisor may rate himself more highly on a self-report, pretest questionnaire designed to measure leadership skills. After completing the course, however, he may be more concerned with impressing the female trainer who taught the course with how much he has learned. As a result, he may adjust his retrospective pretest ratings downward to increase the change accounted for by the post- minus-retrospective pretest scores. At any rate, demand characteristics (e.g., subject bias or social desirability) may cause a shift in a respondent's report from pre- to retrospective pretest, thereby causing a discrepancy.

Secondly, discrepancies between pretests and retrospective pretests may occur when respondents are in different states at the time when the pretest and retrospective pretest are administered. For example, if a client is depressed, he may unconsciously exaggerate the negativity of life situations during the intake interview. Then, if treatment is effective, the bias should be lessened or eliminated at posttest. Aiken and West (1990) refer to this source of bias that occurs in clinical treatment settings as *condition justification*. Aiken and West report that potential clients may unconsciously distort the severity of recent conditions, experiences, or events in order to rationalize their current state of illness or disturbance. The condition justification bias usually leads to overestimates of treatment effects.

A third reason for discrepancies in ratings is that unexplained error, or error that is not systematic, may produce discrepancies between pretests and retrospective pretests. In

this case, the difference is attributed to an unknown cause that may be unique to a particular subject.

A fourth reason for the change that may produce a discrepancy between pretests and retrospective pretests is referred to as the *response-shift phenomenon*. Howard (1980) suggests that a response shift occurs when the experimental intervention changes the subject's evaluation standard with regard to the dimension measured by self-report instruments.

As the discrepancies between pretests and retrospective pretests were discussed by different writers in somewhat ambiguous ways, matters became more confusing. True change, for instance, is referred to as *alpha change* in the organizational development literature (Bedeian, Armenakis, & Gibson, 1980). In discussing the sources of bias in self-reports, Aiken and West (1990) use the term *self-presentation bias* synonymously with what other researchers have referred to as social desirability bias and subject bias. Sprangers (1988) also discusses this same type of bias, but she refers to it as a *response shift*. To complicate matters further, the phenomenon that Howard and colleagues refer to as *response-shift bias* is categorized by Aiken and West (1990) as *experience limitation*.

Of all these, the term *response shift* is of central interest in the literature and is concerned with the change in the meaning, or one's understanding, of the construct of interest. The term appears to be problematic, however, because, as Sprangers (1988) makes clear, several reasons may be given to explain a shift in a respondent's reporting. First, a shift in reporting may be due to what Bedeian and associates (1980) identified as

*gamma change*, which refers to the reconceptualization or redefinition of a referent variable. Although *gamma change* is a unique term, it is an arbitrary term and unfortunately lacks any intuitive meaning. Second, a response shift may be attributed to a change in an individual's internal standard or metric used to evaluate a construct. As noted previously, this change is also referred to as *beta change*, *response-shift bias*, or *experience limitation*. This type of shift in response is concerned with systematic change (i.e., consistent in direction) and is distinguishable from random error. Finally, a respondent's shift in reporting can be due to random error.

As one might expect, a number of investigators became concerned about the potential lack of a common metric and the resulting impact on the use of difference scores. As a result, investigations into the phenomenon known first as *beta change*, and later as *response-shift bias*, began to appear in the literature in the mid- 1970's and continued through the 1980's. The response-shift phenomenon was first addressed by Golembiewski, Billingsley, and Yeager (1976) in the organizational development context. They believed "that an inadequate concept of change leads to diminished or misguided applied research," (p. 133) and therefore it was important to distinguish kinds of change either through experience or by statistical and computational techniques. Their paper defined three types of change: *gamma*, *beta*, and *alpha*.

*Gamma change* refers to the reconceptualization or redefinition of a referent variable and occurs when people change their basic understanding from one testing period to another as, for example, might occur if a person came to view his "boredom" as actually being "depression".

Beta change occurs when a subject's internalized scale of measurement is recalibrated. For example, beta change occurs when, after discounting the occurrence of gamma change, a subject rates a certain behavior as a "3" on a Likert-type scale at Time 1 and the identical behavior as a "4" at Time 2. A clarification of the definition of beta change came from Terbog, Maxwell, and Howard (1982), who noted that measurement of beta change, or response shift, requires that gamma change be ruled out or controlled for.

Finally, alpha change, or behavioral change, is defined as a rating change for which both gamma and beta change have been ruled out (i.e., what is usually regarded as "real" change).

While investigators in the field of organizational development were researching beta change, Howard and Dailey (1979) issued warnings in the educational psychology literature about the validity of self-report measures. They, too, were concerned that it was quite possible that the requirement for a common metric between self-report pre- and post- measures was being violated. Howard, Ralph et al. (1979) therefore, recommended that the conventional pretest-posttest design be extended with a retrospective pretest (referred to by Howard and colleagues as a *thentest*<sup>1</sup>). They claimed that the retrospective pretest, or thentest, is a method by which measures of pre- and posttest levels of functioning are made with the same internal standard of measurement. However, since the retrospective pretest is a method that requires a subjective self-report, its merit was automatically suspect.

---

<sup>1</sup> Howard, Ralph et al. (1979) and colleagues refer to the retrospective pretest as a "thentest". The term, thentest, has since been used synonymously with retrospective pretest.

### **Response-Shift Bias and the Validity of the Retrospective Pretest**

Howard and his colleagues provide a plethora of data to support the use of retrospective pretests to correct for response-shift bias. Initially, Howard, Ralph, Gulanick, Maxwell, Nance, and Gerber (1979) identify a criterion for determining response-shift. To obtain self-report measures of pretest and posttest levels of functioning with respect to the same internal standard, Howard, Ralph, et al. administered a retrospective pretest, their then-test, at the time of posttesting. Subjects were asked to report how they then perceived themselves to have been prior to the intervention. A significant mean difference between pre- and then scores in the experimental condition indicated a response-shift. Since a response-shift is dependent on the treatment, it should occur in the experimental and not in the control condition. If a response-shift is demonstrated, a valid estimate of a treatment effect is given by the mean difference between post- and then- scores.

To investigate the assumptions about response-shift, Howard, Ralph, et al. (1979), conducted five studies involving the evaluation of various treatment interventions. The studies employed an alternative approach to measuring change (i.e., the retrospective pretest-posttest) and compared its accuracy with the traditional pretest-posttest design.

Howard et al. suspect that response-shift bias was responsible for the lack of significant differences between pre- and post- measures and that this confounding effect has influenced a sizeable portion of the literature on program evaluation, counseling outcome, group, attitude, and personality research. Furthermore, they contend that when response-shift occurs, subjects will overestimate their level of functioning when reporting

pretest scores, thereby increasing the probability that the experimental hypothesis will be rejected.

Sprangers (1988) reported that with the exception of a few studies noted by Terborg and Davis (1982), results from numerous studies indicate that a response-shift was found when a retrospective pretest was included. For the most part, mean prescores were higher than the mean retrospective scores. Given this initial overrating, using the mean post- minus-then difference scores resulted in the detection of a positive treatment effect whereas the traditional post-minus-pre difference scores resulted in overtly conservative scores (e.g., Howard & Dailey, 1979; Howard, Dailey, & Gulanick, 1979; Howard, Schmeck, & Bray, 1979; Hoogstraten, 1982; 1985.) Studies conducted by Pohl (1982) and Zwiebel (1987) offer similar conclusions. In a study conducted by Nicholson, Belcastro, & Gold (1985), however, both approaches failed to detect a treatment effect for a university stress counseling program. Although no treatment effect was detected, the authors concluded that the retrospective pretest in lieu of the traditional pretest eliminates a form of response-shift bias that distorts the comparability of pre-posttest measurements. Nicholson, Belcastro, & Gold concluded that the retrospective pretest enables subjects to refer to the same internal construct when completing these measurements and avoids interaction of treatment-by-measurement. Furthermore, they asserted that retrospective pretests would allow empirical analyses of interventions that were begun before a pretest was administered.

A less obvious issue, but certainly a relevant point about these studies, is that the primary finding is concerned with which scores (i.e., post-minus-pre or post-minus-then

scores) yield the greater treatment effects. This conclusion fails to take account of the fact that a number of methodological shortcomings across these studies may make the practice of using retrospective pretests in research premature. For instance, Pohl's study (1982) failed to report essential data (i.e., correlations between post and then test) and was conducted in such a way that may have contaminated the then tests by giving subjects knowledge of their actual posttest performance prior to completion of their retrospective pretest scores. In this study, both the variance decrease and the mean of the retrospective rating indicated that all subjects were less favorable in their post ratings, and this was what is expected when response effects occur.

When the findings began to suggest that the retrospective pretest approach provided a set of conclusions about the effectiveness of an intervention substantially different from that of the traditional pre/post approach, Howard (1980) turned his investigative efforts toward the issue of validity. After conducting five separate analyses of the impact of training interventions (Howard & Dailey, 1979; Howard, Dailey & Gulanick, 1979; Howard, Ralph, et al. 1979) Howard concluded that using retrospective scores provides more valid results than the traditional pre/post scores. This conclusion was based on findings that showed the then/post measurement approach to be more similar to pre/post changes obtained from independent judges' ratings of subject role-playing than were the ones obtained from traditional pre/post self-report methods.

In the early 80's, many investigators believed that the issue of the relative validity of posttest/retrospective pretest versus posttest/pretest difference scores had been empirically established. However, Koele and Hoogstraten (1988) refuted this assumption,

stating that the empirical evidence on this issue was not equivocal (p.51). Although Howard, Millham, Slaten, and O'Donnell (1981) had stated that, in general, thentest/posttest scores are more highly correlated with objective measures of change, Koele and Hoogstraten purported there was evidence to suggest differently. They identified as examples several studies that reported correlations between self-reported change and objective measured change that were not significant, and often, were close to zero (Hoogstraten, 1982, 1985; Sprangers & Hoogstraten, 1987). In the end, however, after implementing their model fitting procedure and reanalyzing Hoogstraten's (1982) data, Koele and Hoogstraten conclude that when a response-shift bias is detected, using thentest/posttest difference scores is the best option. The thentest/posttest difference scores reportedly corrected for systematic nonexperimental effects as estimated in the control group, and they provided an unbiased estimate of the treatment effect.

The literature presented thus far reveals that the retrospective pretest is a potentially useful method for measuring change, and, in most cases, may be superior to the traditional pretest/posttest option when a response shift has occurred. Despite this growing body of literature that supports the potential of retrospective measures, the research community has remained skeptical of retrospective measures. Howard (1980) indicated that this "uneasiness" could be attributed to two distinct factors (p.100). First, he identified a historical/philosophical suspicion of "mentalistic" approaches to research that associates retrospection with introspection, thus creating a negative emotional reaction. And second, problems of response-style bias (e.g., subject compliance, memory distortion, and social desirability) and their effects on the validity of retrospective scores

had not been thoroughly investigated.

### Sources of Bias in Retrospective Pretests

Most research procedures have some problems or trade-offs associated with their use, and the retrospective pretest is no different. In addition to quandaries posed by biases associated with self-reports and the statistical problems associated with analyzing pre-, post-, and retrospective preratings, the retrospective pretest methodology may present problems involving response-style effects (e.g., demand characteristics and motivation), hindsight bias, awareness of an objective measure, possible order effects, and memory distortion (Sprangers & Hoogstraten, 1988c). As a result, a number of studies have focused on identifying how these various sources of bias operate in retrospective pretests.

#### Response-Style Effects

##### *Demand Characteristics*

Although the retrospective pretest method may be helpful in controlling for measurement error associated with response-shift bias, the approach may be more susceptible to subject bias resulting from subject's perception of the demand characteristics of the experimental situation than the Design 4 of Campbell and Stanley (1966). In an attempt to present favorable results, subjects may adjust their retrospective pretest scores in a downward direction. Consequently, a mean pre/retrospective difference score may not reflect a shift in internal standard, but may represent subject bias.

Certainly, one of the most widespread criticism of self-reports of change following treatment has been the operation of social desirability bias and related compliance with implicit task demands to report "improvement" following treatment. Although

posttreatment retrospective self-reports were thought to represent more accurate statements of pretreatment states than pretreatment reports owing to greater familiarity with the behavioral dimensions being studied, it was also possible that such enhanced familiarity with the goals of the intervention might accentuate confounding due to social desirability responding. With this concern in mind, Howard, Millham, Slaten, and O'Donnell (1981) conducted a study to investigate the operation of social desirability and impression management response bias on retrospective measures. A bogus pipeline technique was used in a pilot study to assess the operation of impression management, and results indicated that this technique eliminated the other deceptive component of social desirability and would attenuate subject impression management. Results from the full study indicated that social desirability responding is actually diminished in utilizing the retrospective pretest design. Furthermore, analyses conducted to investigate further the impact of response bias on retrospective measures of change indicated that there is no evidence for the operation of impression management influencing the shifts in evaluation obtained in employing the retrospective methodology. The authors' final conclusions specified that then/post self-report indices of change are more highly correlated with objective and behavioral measures of change than are traditional pre/post self-report indices.

Sprangers and Hoogstraten (1987) implemented a bogus-pipeline induction, and since a response-shift did not occur under bogus-pipeline conditions, it was concluded that response-style effects in the pretesting are a likely basis for response-shift bias. In a replication, Sprangers and Hoogstraten (1988a) examined whether the previous results

were stable and generalizable to a different educational training. Subjects were led to believe that the veracity of the self-reports could be checked by means of objective measures. Contrary to expectation, a bogus-pipeline induction did not lower self-reported preratings. A response shift did not occur in either the bogus-pipeline or the non-bogus-pipeline conditions. The authors speculate that the self-evaluations at pretest could not be improved by removing potential response-style effects because such effects were simply not operative. Since it is quite possible that the subjects had some familiarity with aspects of the First Aid training that was taught, their self-evaluations may have been valid. The replication demonstrated that a construct not susceptible to removal of response-style effects is not susceptible to response-shift bias either.

An alternative method for dealing with the problem of demand characteristics is to include a placebo control condition. Since placebo subjects devote the same amount of time and effort to the placebo treatment as do experimental subjects to the experimental intervention, a significant mean pre/then difference score in the placebo condition can be attributed to subject bias, thus invalidating the response-shift interpretation. In addition, the design should include a no-treatment control condition to control for rival hypotheses. To rule out subject bias as a viable alternative explanation of the response-shift interpretation, the mean pre/then difference score should not be significant in either placebo or no-treatment control conditions.

Sprangers (1989) was concerned that subject bias was not controlled for in many studies employing retrospective designs that also demonstrated a response shift. Sprangers observed that in some cases, designs including a retrospective pretest (thentest)

may be more susceptible to subject bias resulting from subjects' perception of the demand characteristics of the experimental situation. Since the thentest is administered at the same time as the posttest, subjects may adjust their thenscores in a downward direction in an attempt to indicate favorable results. Consequently, a mean pre/then difference score may not reflect a shift in internal standard, but may represent subject bias instead. In order to investigate whether subject bias was an alternative explanation of the response-shift interpretation in reported studies, Sprangers conducted a review of the research. To summarize, she did a computer search for articles that included the Howard, Ralph, et al. (1979) article in their reference list. Of the 24 studies, 16 did not warrant any statement about the occurrence of a response-shift bias, relative to a subject bias because they lacked a placebo control condition. In the eight studies that did include a placebo control condition, only one study reported significant mean pre/then difference scores found in both experimental and placebo control conditions, thus supporting the hypothesis of subject bias. Sprangers notes three supplementary considerations:

In one of the studies conducted by Sprangers and Hoogstraten (1988b), they found significant mean post/then difference scores in both the experimental and placebo control conditions, and not in the no-treatment control condition. They concluded that mean post/then difference scores were not free from subject bias.

In additional studies where the order of the thentest and posttest were reversed, results suggest that neither post- nor then-scores were affected. The authors of these studies concluded that the thentest seems rather robust against order manipulations (Sprangers & Hoogstraten, 1987; Terborg & Davis, 1982).

Third, using an alternative procedure, Howard et al. (1981) investigated the operation of subject's response-style effects on thenscores. A prevention or reduction of response-style effects was planned, utilizing a bogus-pipeline technique at the posttesting. The hypothesis of greater bias in thenscores was not supported. Sprangers and Hoogstraten (1987, 1988a), who induced a bogus-

pipeline procedure at pretesting, also found no support for subject's response-style effects in the scores (Sprangers, 1989, p. 14).

Taking these supplementary considerations into account, Sprangers concludes that subject bias has not proven to be a consistent alternative explanation of the response-shift interpretation in the reported studies. In other words, since subject bias has been demonstrated, researchers should be aware of this form of bias in future research. If, for instance, a researcher suspects that the research question will lend itself to promoting subject bias or social desirability, then a control condition should be added in order to differentiate between response shift and subject bias.

Although it is apparent that including a control group as part of the retrospective pretest will help to rule-out or at least reduce threats to internal and external validity, there may be occasions when a suitable control group is not available or is too costly. An alternative would be to find an appropriate comparison (control) group. Toedter, Lasker & Campbell (1990) extended the retrospective pretest approach by including a surrogate comparison group in a longitudinal investigation of bereavement outcome following pregnancy loss. Based on this study, the authors concluded that the use of the retrospective pretest with a comparison group has potential as a quasi-experimental approach to the study of certain specialized populations. Furthermore, they indicate that the value of this approach "lies in enabling the researcher to gather pretest data in populations in which the gathering of such data in a prospective fashion would be impossible (as in the case of a trauma such as rape or miscarriage) or prohibitively expensive (as in the case of stillbirth or neonatal loss)" (p. 87). This design also provided

the authors with a means of checking the accuracy of retrospective data that they had reason to suspect had been distorted by the experience of a subsequent event.

Following-up on the concern about subject bias in self-report measures, Sprangers and Hoogstraten (1991) investigated the influence of subject bias in self-reported post-minus-pre difference scores, post-minus-then difference scores, and direct improvement scores. The study design included an experimental, a placebo, and a no-treatment control condition. The 64 psychology freshmen who were enrolled in either a text-studying training or a study-planning training also completed a performance task at pre- and posttraining sessions. Results indicated the effectiveness of both study skills training. The major conclusion, however, is that the three measures of self-reported change were not free from subject bias. Based on the implications of the findings, the authors offer the following four recommendations to social scientists who employ self-report measures to assess improvement programs (p. 12). First, be alert to the susceptibility of self-report measures of change to subject bias, especially in circumstance in which subjects voluntarily exert effort to improve their skills. Second, each of the measures has its own limitation and advantages. However, whenever pre- and post- data are collected, it is advisable to collect *then* data as well. Post-minus-pre difference scores may be affected by response-shift bias, and if this occurs, post-minus-then difference scores are to be preferred. Third, the authors dispute the quest for formal evaluation (i.e., performance measures) to assess improvement programs (e.g., Conway & Ross, 1984). They contend that the measurement of change is replete with problems whether one uses self-reports or objective measures. Finally, the authors state that social scientist should anticipate that self-report

measures may produce biased results and therefore include both performance and self-report measures in their studies. In addition, a placebo and no-treatment control group in addition to the experimental group will enable one to control for subject bias.

### *Motivation*

There is no formal research investigating how a person's degree of motivation may affect retrospective recall and contribute to measurement error, however, it is an important question that warrants further research.

In 1978, Rippey, Geller, and King took an initial step in testing the validity of the retrospective pretest. They used the retrospective pretest to evaluate gains derived by medical students from a summer course that was intended to help them prepare for their Board of Medical Examiners tests. The authors made some interesting observations regarding motivation. They suggest that, to employ a retrospective test successfully, one must have subjects' cooperation. For instance, medical students may be resistant to the idea of taking a test twice, once "for real" and once "for pretend" (p.488). Or, those who were uncooperative or antagonistic toward evaluation might produce useless material whereas students who were interested and cooperative would be likely to produce valid scores from which prior states of knowledge could be inferred. The authors concluded that subjects' motivations play an important role in determining the validity of the retrospective pretest approach used to evaluate an educational program.

Just as response-style effects such as social desirability play a role in affecting conventional pretraining scores, there is evidence that suggests that motivation plays a similar role. For instance, Sprangers (1988) discovered that when subjects are reluctant or

less motivated to provide valid data at pretest, a response shift may occur. When evaluating two text-studying training programs, both were effective, but only one indicated that a response shift had occurred. Sprangers found that the response shift occurred in the text-studying program that contained students whose participation was required. The other program contained students who voluntarily assigned themselves. Apparently they considered themselves in need of such a training, and, therefore, may have toned down their self-assessments, thereby providing a more accurate pretest rating. This was evidenced by the difference between the mean pretest scores of training participants and their control counterparts whose participation was required (p. 105).

Another example of how motivation may play a role in altering retrospective accounts was brought to bear by Catania, Gibson, Chitwood, and Coates (1990), who were conducting AIDS research. The researchers observed that highly motivated respondents may try harder to understand and answer questions, and less motivated subjects may skip items or give less thoughtful answers. For instance, when subjects are required to exert momentous effort to provide a retrospective account of a complex event, differences in motivation will most certainly impact measurement error.

#### Awareness of an Objective Measure

As investigators became aware of response-style effects in retrospective pretests, they began to question whether a respondent's awareness of an objective measure helps to reduce the response-style effects in retrospective pretests as it does in pretests. This was due, in part, to a common assumption among researchers that self-report data are more accurate when a behavioral measure or an observed variable are included. This

assumption is supported by the common practice of validating self-report measures by comparing them with behavioral measures of the same dimension of interest (Howard, Maxwell, Wiener, Boynton, & Rooney, 1980). In addition, researchers believe that the validity of subjective measures is raised by the awareness of an objective measure. For instance, in a study conducted by Bauman and Dent (1982), the self-reports of adolescents who had smoked recently showed more smoking if these persons were informed that an objective measure was to be obtained as well. While using behavioral measures to validate self-reports may be appropriate in general psychological research, this practice may have different implications when used with retrospective self-reports.

As a result, a number of investigators have tried to determine whether the validity of self-reported behavior would be influenced by knowing that an objective measure of the same behavior was also to be obtained. Howard, Millham, Slaten and O'Donnel (1981) found that the greater the familiarity with the behavioral dimensions being studied, the better the opportunity for valid self-evaluations. Howard and his colleagues' work demonstrates that then/post comparisons were more consistent with independently obtained behavioral measures than were pre/post comparisons. Howard et al. also found that the test/posttest scores are more highly correlated with objective measures of change than are pretest/posttest scores.

Hoogstraten (1985) conducted two experiments that included retrospective preratings to assess the influence on subjects' self-ratings of awareness of an objective measure and taking a performance test. While a response shift was demonstrated in both experiments, Hoogstraten found that awareness of an objective measure tends to weaken

the response-shift phenomenon, whereas actually taking a performance test tends to strengthen it for knowledge ratings only. These findings were consistent with results set forth by Howard, Schmeck, and Bray (1979), which indicated that subjects who are asked to take a performance test to demonstrate their knowledge of the course material may be influenced on subsequent retrospective self-ratings of performance.

Before leaving this discussion on objective measures, one last study involving retrospective data should be considered. Crockett, Schulenberg, and Petersen (1987) examined the validity of self-report data in a sample of young adolescents for whom objective and self-report data were available on course grades, height, and weight. Although this study did not incorporate a retrospective pretest, it required adolescents to provide retrospective accounts of grades from previous years. In order to minimize errors of recall, students were asked about their typical (i.e., modal) grades rather than being asked to recall grades in specific courses. The correlations for course grades were uniformly high, with coefficients ranging from .70 to .84. Although the results suggest that the self-reports were fairly reliable, the correspondence between self-reports and objective data was not perfect. In addition, there was evidence of systematic error. For example, boys tended to overestimate their course grades, height, and weight while girls' reports were underestimated. The important lesson to be gleaned from this study is that when choosing self-report methods, researchers must determine how much error they can tolerate. If substantial precision is required, then an attempt should be made to obtain objective measurements. If objective measurement is not possible, it would appear that retrospective, self-report data can be a useful alternative.

### Hindsight Bias

An additional biases that has the potential to affect recall is hindsight bias.

Hindsight bias is the tendency for people considering a past event to overestimate their likelihood of having predicted its occurrence. For example, after the University of Arizona Wildcats basketball team won the NCAA championship in 1997, numerous “fans” claimed to have known that the Cats would win this national contest. This is highly unlikely given that the odds of the Wildcats making it through the first draft were extremely low. Fischhoff (1975) has noted that hindsight bias has been demonstrated in a variety of applied settings including politics, historical judgment, psychotherapy case histories, medical diagnoses and employee evaluations. Fischhoff contends that the most important implications about hindsight bias is that it makes it difficult to learn the lessons of the past.

A number of studies have investigated the hindsight phenomenon. The literature includes studies aimed at discovering variables that influence the magnitude of the hindsight effect (Christensen-Szalanski & Willham, 1991). Mark and Mellor (1991), for example, examined the extent to which a self-relevant, negative outcome effects hindsight bias. In other studies, psychologists investigating theories of attribution have examined the differences in motive strength in a hindsight bias paradigms. As an example, Campbell and Tesser (1983) found that individuals higher in dogmatism or in intolerance for ambiguity showed more hindsight bias.

### Order Effects

Armenakis, Buckely, and Bedeian (1985) suggested that a better understanding of

the measurement of change might be accomplished through more research on the time dimension. In response to this suggestion, Sprangers and Hoogstraten (1988c) investigated how time affects retrospective preratings. They proposed that there may be instances in which the administration of the retrospective pretest is delayed. In some studies, it may be advantageous to examine a delayed or lasting treatment effect by administering the post- and retrospective pretest some time after the experimental intervention. In other cases, for practical reasons, it may not be possible to administer a post- and retrospective pretest immediately after completion of the experimental intervention. In this study, they examined to what extent retrospective preratings are affected by a follow-up assessment, two weeks after posttesting, and a delay of the assessment, two weeks after treatment. The study design consisted of an experimental, placebo, and a no-treatment control condition. The results indicate that, for the psychology freshman fulfilling a course requirement, a 2-week time interval does not exert an influence on readministered nor on delayed retrospective preratings.<sup>2</sup>

To address the question of whether administration of the retrospective pretest independent of the posttest affects the ratings, Sprangers and Hoogstraten (1989) reversed the order of the administration. They found that reversing the order of the posttest and retrospective pretest did not affect the self-report posttest scores or the retrospective pretest scores. They concluded that the order of administering the posttest and retrospective pretest seems arbitrary.

---

<sup>2</sup>The Howard et al. study (1979, Study 3) demonstrated comparable results; similarities were found between the 1-year follow-up retrospective prescores and the initial retrospective prescores.

### Memory Distortion in Retrospective Pretest

Since the introduction of the retrospective pretest report, there has been concern among researchers about the error associated with the memory distortion (i.e., the reliability of recall) that may occur as part of the retrospective process. If, as Kotre (1995) suggests, the remembered past is a creation of the present, then how reliable can retrospective reports be? On the other hand, isn't every self-report retrospective in nature? There are, no easy answers to these questions, but the literature does provide some insight about the conditions under which retrospective pretests can be used with confidence to minimize distortion.

It should be noted, however, that the term, *memory distortion*, appears to be a "catch-all" label used to describe a form of systematic error accounting for discrepancies found between pretest and retrospective pretest data. The term also lacks consensus among investigators as to the criteria that defines memory distortion in retrospective pretests. As a result, numerous conclusions about memory distortion are found in the literature, but many of these claims could be categorized as forms of response-style effects.

Concern about retrospective measures being susceptible to memory distortion has existed almost as long as the retrospective pretest. Formal discussions in the literature began when Campbell and Stanley (1963) initially wrote that, due to memory distortion, pretests are more accurate than retrospective pretests and that retrospective pretest data should be collected only when pretest results are unavailable. However, Campbell and Stanley (1963) maintained that the probable direction of memory bias is to distort past

attitudes into agreement with present ones or into agreement with what is believed to be socially desirable attitudes. Thus, memory distortion seemed more likely to disguise a significant treatment effect rather than masquerade as one. Furthermore, Campbell and Stanley suggested that retrospective pretests can be used as partial curbs to rival hypotheses of history, selective mortality, and shifts in initial selection (p. 66).

Howard noted that in a personal communication with Campbell about retrospective measures being susceptible to memory distortions, Campbell stated that, "this conclusion was drawn without consideration of the possibility of an instrumentation with experimental treatment confounding for self-reports and should be considered if further research reveals that retrospective self-report pretests provide results which are more accurate than the usual self-report pretests" (Howard, Ralph, et al., 1979c, p.6). In other words, Campbell acknowledged that memory distortion was a phenomenon that required further investigation to define the conditions in which this systematic bias operates in retrospective data.

To further investigate this the role of memory distortion, Howard, Ralph, et al. (1979), designed Study 5 with a twofold purpose: 1) to determine the validity of then/post measures relative to pre/post self-report indices of change, and 2) to ascertain if pre/then differences (attributed to response-shift in the previous four studies) might be due to systematic memory distortion. As hypothesized, the then ratings were significantly lower than either mean pre or mean memory ratings. (Memory ratings were collected after subjects completed the then/post questionnaire. The memory questionnaire instructed subjects to record what they remembered their precourse ratings to be.) In

addition, subjects' thentest scores were consistently lower than their actual pretest estimate and also their memory of those pretest ratings. Since no differences between pretest and recalled pretest scores were found, and biases or inadequate memory of the actual pretests ratings do not explain these results, Howard et al. concluded that the established conventional pretest-retrospective pretest difference was due to a response shift. As a result of this study and one conducted by Howard and Dailey (1979), memory distortion was considered to be an alternative explanation of the response-shift phenomenon.

Researchers differ in the implications they have attached to the incompatibility of pretreatment ratings, and this in turn affects what is classified as memory distortion. For instance, contrary to Howard, Ralph, et al. (1979), Collins Graham, Hansen, and Johnson (1985) suggested that memory distortion in the experimental condition was a sign of the response-shift phenomenon. They contend that if a treatment produces a response-shift bias, treatment subjects will use the response scale differently at posttest than will control subjects. When treatment subjects are asked to recall pretest scores, they will use a response scale different from the one on which they based their pretest report. As a result of response-shift bias, the mean difference between pretest and recalled pretest scores should vary by experimental condition. Collins, et al. therefore consider memory distortion in the experimental condition as a sign of the response-shift phenomenon.

Although Sprangers and Hoogstraten investigated these contradictory views of memory distortion, their data were not conclusive. They found results about memory distortion that differed from both Howard, Ralph, et al. (1979) and Collins et al. (1985).

In a study conducted by Sprangers and Hoogstraten (1989, Experiment 1), in which there was no treatment effect and a response shift failed to occur, experimental subjects could not recall their pretreatment ratings. Sprangers and Hoogstraten suggest that, “These inconsistent results do not warrant firm conclusions about the viability of memory distortion as an alternative explanation or concomitant phenomenon of the response shift” (p. 271).

Research efforts to identify and clarify the role of memory distortion failed to minimize concern about memory distortion, and as a result, criticisms of retrospective pretest persisted in the literature. Armenakis, Buckley, and Bedeian (1986), for instance, criticized the retrospective pretest in the organizational literature based on the assumption that respondents are incapable of recollecting accurately and articulating past events. In discussing this finding, Sprangers (1988) regards these conclusions as inappropriate because in the experiment cited, the Armenakis et al., focused on perceptions of phenomena external to respondents or on factors external to their behavior which are not as salient. To support this contention, Sprangers (1988) cites Porras and Singh who found that personal experiences are more salient and as a consequence more likely to be recalled accurately. More importantly, Sprangers notes that Armenakis et al. focus primarily on recollection of self-reported ratings provided at an earlier time, and this is asking respondents to perform a recall task. The retrospective pretest, however, does not ask respondents to recall their earlier rating, but rather to recall the conditions that existed before the training began, and to assess these conditions again from a potentially new, recalibrated perspective (Sprangers & Hoogstraten, 1989). In conclusion, since neither of

the mentioned studies employs a retrospective measure that taps personal experience, they appear to describe situations different from the ones encountered in the studies by Sprangers and Hoogstraten and Howard et al.

Additional studies have reported that memory cannot be relied on to produce accurate recollections (Rippey et al. 1978, Armenakis & Bedeian, 1982). In response to these allegations about the unreliability of recall in retrospective pretests, Sprangers (1988) reports that she and her colleagues, De Meijer and Hoogstraten have argued that each self-report is retrospective in nature, in that it is a self-appraisal based on past experiences. In their view, there is no distinction between conventional and retrospective preratings (p.99).

An interesting study that supports the perspective expressed by Sprangers, De Meijer and Hoogstraten involved self-reports of services by homeless mentally ill clients. Calsyn, Allen, Morse, Smith and Tempelhoff (1993) found that the reliability and validity of self-report data provided by homeless mentally ill clients were generally favorable. Although a retrospective pretest was not used in this study, clients' were asked to provide self-reports of service use for the previous month. Clients' reports agreed with treatment staff estimates for most service categories, thus demonstrating the reliability of recall for this group. Consequently, if self-reports based on retrospective accounts from mentally ill clients yield valid data, then maybe we should not worry so much about the validity of retrospective pretests.

### *Autobiographical Recall*

Concerns about the reliability of recall are also documented in the memory

literature. Autobiographical recall, for instance, is potentially biased in that it may facilitate a unified, positive view of self. Conway (1990) and his colleagues previously demonstrated that what people remember from their pasts may illustrate both their personal histories and their current attitudes or beliefs. People may therefore recall past attitudes and behavior in a manner that increases their consistency with their current attitudes. In order to determine whether people recall their behaviors and expectations in ways that increase their consistency with subsequent outcomes, Conway had Canadian students report their preparations and expectations for an exam in an undergraduate course. Weeks later, after they had received their grades, they were unexpectedly asked to recall as accurately as possible their earlier reports. Those whose actual grades were worse than anticipated recalled working less and expecting less satisfaction relative to their initial reports. Students whose actual grades were better than anticipated recalled expecting higher satisfaction relative to their initial reports. Since similar results were obtained in a replication study and over 90% of the student in both studies did not receive their expected grades, Conway (1990) concluded that the findings support the view that autobiographical recall may be generally reconstructive and potentially biased. Although Conway refers to this as “recall bias”, it should be noted that he also indicates the bias allowed students to maintain a positive self-view of the self. Therefore, this recall bias could be classified as a form of self-presentation bias rather than a form of memory distortion. One should also note that this is a study where self-presentation bias affected the retrospective pretest rather than the pretest.

### *Complex Episodes*

The accuracy of the retrospective recall involved in complex episodes has also been investigated. Research with eyewitnesses to crimes, for example, has provided ample evidence of a witness's tendency to fill in descriptions, dates, locations, and duration estimates (Buckhout, 1974; Loftus, 1979). For example, when a moderate earthquake struck the general region of Westchester County, New York, Buckhout, Fox, and Rabinowitz (1989) used this event to investigate the accuracy of retrospective recall. Within a few days following the minor earthquake, the researchers collected data from 246 people on their retrospective recall of the duration of the earthquake, along with estimates of arousal, alertness, and attention to media. Previous research led the authors to suspect that duration estimates of the earthquake would be longer than the actual duration of the quake because the subjects had no choice but to be passive observers of an arousing, nonuniform, and novel situation that they could not control. As anticipated, findings indicated that the duration of the quake was pervasively overestimated, with higher overestimates from people in the area where the most intense quake occurred.

### *Personal Attributes*

Studies investigating the recall of personal attributes indicate that the very request influences the task of recall (Catania, Gibson, Chitwood, and Coates, 1990). For instance, if you ask subjects to recall how assertive they were at work prior to an assertiveness training course, most subjects would not find this request to be too personal (i.e., obtrusive or of a sensitive nature). However, if subjects are asked about their levels of drug use and sexual activity as part of an AIDS intervention, they may be less willing or

able to generate accurate information. Eliciting responses about “sensitive” or personal issues may certainly increase measurement error by altering respondents motivation or through self-presentation bias. Hence, the task of recall, whether it is pre-, post-, or retrospective, is influenced by the very nature of the respondent variables.

In a report on methodological problems in AIDS behavioral research, Catania, Gibson, Chitwood, and Coates (1990) identify four memory-related issues that affect respondent influences on measurement error. The first two issues, *vividness and complexity* of the material to be recalled, is influenced by the length of the recall period, the vividness of the events to be recalled, the difficulty involved in recalling past events, and the respondents’ motivation for recalling the past experiences. Thirdly, *inferential recall strategies* are used to estimate frequencies of behaviors or prevalence of attitudes. Bradburn, Rips, & Shevell (1987) found that decomposing and the availability heuristic have been used to estimate frequencies of an occurrence. Decomposition results when respondents derive a rate of occurrence for a given time period and then multiply that rate to arrive at some total figure for a larger period of time. Decomposition works best for subjects who have routine patterns of activity and infrequent occurrences; for people with many occurrences, decomposing may prove problematic. Finally, emotional issues or the personal salience of an event may influence how vividly that event is recalled. Positive and negative emotions associated with an event can influence recall. For instance, events associated with highly positive emotions are more likely to be recalled than events coupled with negative feelings. However, how emotional states at the time of interview influence the degree and direction of measurement error requires further investigation.

### *Bias Associated with Variable Types*

Memory distortion has also been used as an explanation for the inaccuracies found in the recall of the diverse variables utilized in retrospective pretests. Following a review of the retrospective pretest literature, it is clear that some variable types are expected to be more prone to memory distortion than other types. The term *variable types* is used to describe the type of data that is being gathered in a retrospective pretest. Some of the more common types of retrospective variables include attitudinal, cognitive, experiential, and opinion-oriented. According to the literature, a variable type can address a wide range of topics. Sprangers (1988) identified numerous studies that contain attitudinally oriented variables including: assertiveness, Christian growth, communication skills, disability awareness, organizational change, and teacher training.

It is not uncommon for retrospective pretest studies to include one or more questions about the variable types that are included. Most often, questions about variable types will focus on identifying which variable types are more prone to a response-shift, or examining which variable types work most effectively in retrospective pretests. Although a review of the literature reveals that these questions are asked most often about attitudinal and cognitive variables, a third area of inquiry addresses variable complexity.

### *Attitudinal Variables*

The few studies addressing attitudinal and cognitive variables in retrospective pretests have primarily focused on identifying which variable is most subject to a response-shift. Intuitively, it would seem that studies evaluating training programs that include attitudinal variables would be more subject to response-shift effects than studies

addressing cognitive variables. However, the literature suggests that this is not the case. Pohl's (1982) study of 51 students enrolled in a three-week study of probability theory measured cognitive variables only. Pohl expected that cognitive variables would be less subject to response-shift than attitudinal variables. Nevertheless, a significant response-shift was observed. Pohl concluded, however, that response-shift may be even more dramatic when involving attitudinal variables. Numerous retrospective pretests studies have included both attitudinal and cognitive measures to evaluate training interventions, and although differences are observed, there is no conclusive evidence to suggest that attitudinal variables are more subject to response-shift than cognitive variables.

A study conducted Zwiebel (1987) examined the effectiveness of attitudinal variables in retrospective pretests. Zwiebel compared two techniques, the traditional "pre/post" to the "then/post" approach, in measuring change in attitudes toward disabled individuals. Students who attended a special education course designed to develop their emotional acceptance and empathy toward mental retardation were given attitude measures before, after, and retrospectively. Comparison of the experimental and control groups using the two measurement techniques show that change was found only by using the retrospective pretest approach. The author concluded that since attitudinal ratings may be influenced by psychological stimuli, change can only be measured by comparing one's feelings at different times on the basis of evaluations made in the same psychological or cognitive state. Given that the course may alter the subject's cognitive structure, thereby affecting self-ratings, the retrospective pretest is the best approach for capturing attitudinal change.

An additional problem associated with the use of attitudinal variables in retrospective pretests relates to the reciprocal relation between attitudes and recall of attitudinally relevant past behaviors. Ross, McFarland, Conway, and Zanna (1983) conducted three experiments to investigate this relationship. Although this study was not based on a retrospective pretest approach, the findings provide relevant data about the relationship of attitudes and recall, both of which play an important role in providing information about the validity of retrospective responses. In previous studies, Ross and associates (1983) found that people selectively recall their past behaviors to make them consistent with their current attitudes. In the Ross et al. study, all three experiments provide support for the reciprocal relation between attitudes and recall; that is, attitudes can affect recall, which will in turn affect commitment to the attitudes.

Before leaving the discussion on attitudinal variables, it is important to note that the term *attitudinal* applies to a continuum of attitude-based variables. Some references to this term are based on a dictionary definition, and that is, “a mental condition or feeling with regard to a state or fact” such as a mood state (Merriam-Webster Dictionary). Other references to attitudinal variables describe a disposition or a trait dependent characteristic (e.g., optimism). Obviously, states will be more likely to change than traits, and this aspect may affect the reliability of recall in the attitudinal variables used in retrospective pretests.

### *Cognitive Variables*

Cognitively oriented variables have been used in studies on psychotherapy, counseling, drug abuse prevention, learning theory, parent-child interactions, probability

learning, problem solving strategies, and study skills (Sprangers, 1988, p.104). In some studies, the term *cognitive* literally refers to knowledge. For example, in Pohl's (1982) study, students enrolled in a three-week study of probability theory were asked to recall how much they knew about probability theory. In other cases, the term cognitive variable may be referring to perceptions or intuitions. An intuition may be more difficult to recall because it is the "faculty of knowing things without conscious reasoning" (Merriam-Webster Dictionary). Most of the questions asked the use of cognitive variables in retrospective pretests looked at how prone these variables are to a response-shift.

In an experiment investigating the effect of taking a performance test on self-reports, for example, Hoogstraten (1985) decided to test the notion that cognitive variables may be less subject to response shift than attitudinal variables. He employed eight attitude-oriented variables and seven cognition-oriented variables, but the findings did not support the hypothesis. Hoogstraten found that taking a performance test did lower retrospective pretest ratings of cognitive variables. As a result, only the cognitive variables were associated with a response shift.

Rippey, Geller, and King (1978) examined the validity of retrospective pretesting in the cognitive domain. This study examined the effectiveness (i.e, accuracy) of cognitive variables. A retrospective pretest was used to evaluate a summer course for American students attending foreign medical schools who were seeking to transfer to school in the U.S. After completing the course, students were asked to answer items on a test twice: once with the response they thought correct and once as they thought they would have responded before the summer course. Findings indicated that over half of the group (i.e.,

20 of the 34 students) had at least 50% accuracy of recall (range: 50-78%) of their precourse responses. Given the recall accuracy of these cognitive variables, the authors concluded that the retrospective pretest may be a useful procedure in evaluating the gains derived from an educational program, especially when the opportunity to determine a baseline of information is limited or absent.

#### *Variable Complexity*

In addition to attitudinal and cognitive variables, retrospective pretest data has included an array of variable types including physical and mental health states or traits, psychological treatment, and social support. Toedter, Lasker, and Campbell (1990), as noted previously, investigated bereavement outcome following pregnancy loss. Since the question of mental status (e.g., anxiety) is sometimes in the study of pregnancy loss, mental health and social support measures were employed. The authors chose these variable types because they thought there was reason to believe that it is easier to recollect specific symptoms than more complex material such as feelings and cognitions. This comment raises an important question. After taking into account the previously identified sources of bias in retrospective pretests, which variable types contain the least amount of memory distortion? The current literature provides hints to this answer to this question. For example, it is not uncommon to find authors who suggest that physical symptoms or data that include reports on health are easier to recall and probably more reliable than attitudinal or cognitive data. Or, a recent study by Southwick, Morgan, Nicolaou, and Charney (1997) reported that some researchers have described traumatic memory as fixed and indelible. However in a prospective investigation of memory for serious combat-

related traumatic events, the investigators found that one month after the Gulf War, 46% of subjects reported one or more traumatic events that they did not recall 2 years later. Furthermore, 70% of subjects at the 2-year evaluation recalled traumatic events that they had not reported at 1 month (p. 175). Although the investigators offer several psychological explanations for the subjects' discrepant reports, we do not know how much of the distortion can be attributed to the complexity of the variables subjects were asked to recall.

Our understanding of memory distortion as a source of bias in retrospective recall could be enhanced by organizing retrospective data according to the level of variable complexity. Figure 1 provides a heuristic aide for understanding this approach. As the level of variable complexity increases from left to right, so too does the difficulty of the recall task. In addition, the first three event categories on the left side of the figure are also the types of variables that easiest to verify or to use with objective measures. A patient's report of visits to the doctor, for example, can be verified through patient records. Or, even traumatic events can be verified, to some extent, through eye witness reports.

Each level of complexity could be also classified on a dimension of time (e.g., weeks, months, or years), and the percentage of accurate recall should be recorded for a particular recall task. The approach is much like assembling a puzzle; pieces of data are placed along the continuum of variables, and someday the *modus operandi* of memory distortion will be more clearly understood.

The selection of the variables that are to be included in the retrospective pretests

warrant further investigation since it appears to be a central issue in the use of this approach. As Toedter et al. (1990) acknowledged, the psychometric characteristics of measures play an important role in contributing to the clear picture that emerges from retrospective accounts. The careful selection of the variables used in retrospection may help to reduce memory distortion and rule out rival plausible hypotheses.

**Table 1: Memory distortion in retrospective data across time.**

<b>Specific Events</b>	<b>Complex Events</b>	<b>Traumatic Events</b>	<b>Opinions (Beliefs)</b>	<b>Attitudes (States)</b>	<b>Attitudes (Traits)</b>	<b>Cognitions (Intuitions)</b>
behaviors symptoms services	experiences series-of - details	shocking news flash memories	evaluations judgements	feelings moods	dispositions health status	knowledge perceptions

### Summary

The review of the literature illustrates that investigators from various fields including education, psychology, and health have questioned the validity of the retrospective pretest since its formal introduction. Researchers and evaluators alike are concerned about the reliability and validity of this approach when it is used to measure change following a training program or intervention. Researchers are concerned that, like self-report measures, the retrospective pretest may be affected by subject bias, response-style effects, and memory distortion (Sprangers, 1988). Previous studies have examined the effects of subject bias and response-style effects, but less attention has been given to the memory distortion associated with retrospective recall.

While the memory literature addresses the reliability of recall and sources of measurement error in the recall autobiographical data, complex episodes, and personal

attributes, this body of literature lacks methodological studies that systematically identify sources of error in the different types of variables employed in the retrospective pretests approach. Therefore, a more systematic method of research is needed since the psychometric characteristics of measures play an important role in contributing to the clear picture that emerges from retrospective accounts. The goal of systematic approach should be to understand, to what degree and in what direction, measurement error varies across these categories of variables, and to determine what impact that length of time may have on the reliability of the retrospective responses.

#### Present Study

The primary goal of this research project was to develop a methodology to measure the systematic error (or memory distortion) that may be associated with variables that involve a range of recall tasks (e.g., behaviors, opinions, attitudes, and cognitions). The research project included two phases. The purpose of Phase 1 was to confirm the hypothesized variable types or factors on an independent sample. In addition, these factors were cross-validated on a second sample of students.

Phase 2 seeks to identify variable types and individual survey items that are subject to memory distortion. To assist in the determination of what effect the variable types have upon pretest and retrospective pretest scores across three points in time, Phase 2 will apply structural equations modeling (SEM). Phase 2, however, does not include a formal intervention. Omitting a formal intervention avoids any treatment by instrumentation effects; therefore, memory distortion may be easier to assess. In order to capture naturally occurring changes in students behaviors, opinions, and attitudes over the course of the

semester, the content of the retrospective survey focuses on the student's school experience. Based on the lack of systematic investigation and methodological support in this type of research study, a number of research questions have been formulated to guide the investigator.

### Research Questions

1. Are there significant differences in the magnitude and the direction of change in the pretest and retrospective pretest scores for the individual survey items at different time points?
2. Are there significant differences in the magnitude and the direction of change in the pretest and retrospective pretest scores for the identified variable types at different time points? (For example, are retrospective reports of attitudes more reliable than retrospective accounts of cognitions?)
3. Is there less agreement between pretest and retrospective pretest scores for the items that require "self-ratings" as opposed to the items that require a "self-report" response? (Since ratings of personal attributes (e.g., academic ability) are more susceptible to subject bias, it is hypothesized that self-ratings will be less accurate than general self-reports.)
4. Are there significant differences in the magnitude and the direction of change in the pretest and retrospective pretest scores for those items identified as being susceptible to a response shift? (A response shift is identified by either an upward or downward shift in the direction of retrospective pretest scores).
5. Are there significant differences in the magnitude and the direction of change in the pretest and retrospective pretest scores for items identified as being susceptible to

hindsight bias? (If hindsight is operational, there should be a high correlation between posttest and retrospective pretest scores.)

6. Are attitudinal variables more subject to a response shift than opinion or cognitive variables?

7. What amount of variance in a retrospective pretest can be predicted by variable type having taken into account the pretest score? This question will be examined in the analytical model shown in Figure 1.

The structural equations model will examine the individual patterns of change in students self-reports. In this model, the path from Pretest to Posttest represents stability. Since there is no formal intervention, what a student is like at pretest should not change much at posttest; therefore, a high path coefficient represents stability in the variable type that is measured.

The path from Pretest to the Retrospective Pretest provides a reliability check (i.e., test/retest). Essentially, if memory were perfect, what is reported at pretest should be the same in the retrospective pretest, unless, of course, there is a theory to explain why memory could be altered in the recall of a particular variable type (e.g., response shift).

The path from Posttest to the Retrospective Pretest identifies how much memory reconstruction has occurred. At posttest, the retrospective state is influenced by the present state; therefore, students will describe what they were like in the past as they see it now.

The paths leading from the variables Time to Posttest and Time to Retrospective Pretest identify the effect of forgetting. In other words, these paths identify the effect of

longer intervals of time on both the posttest and retrospective pretest. Finally, the double-headed arrow between Pretest and Time tests for the possibility of bias in group selections.

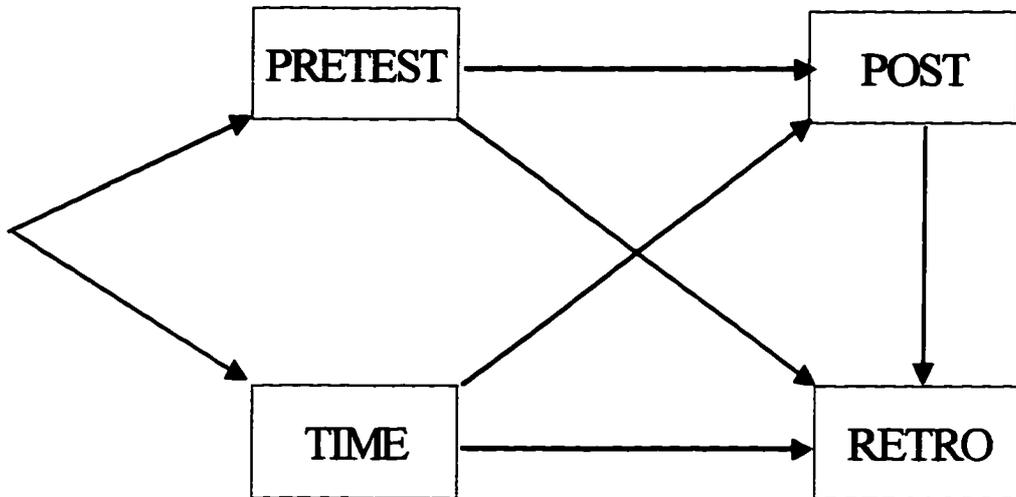


Figure 1: Analytical model used to test for memory distortion in retrospective pretests.

## CHAPTER 2: PHASE 1: IDENTIFICATION AND CROSS-VALIDATION OF VARIABLE TYPES

Phase 1 is designed to confirm and cross-validate the variable types. The confirmation process involves two surveys, the New Student Survey (NSS) and the Beck Depression Inventory (BDI). The New Student Survey (NSS) was developed for use as a pretest instrument. Differences in students' reports on the NSS and a retrospective version, the Retro Student Survey (RSS) were used to identify levels of memory distortion. The RSS includes the same items as the pretest version, but items were written in the past tense form. The BDI, on the other hand, is a well-accepted measure of general depression and is employed in this study to evaluate changes in students' mood states. A retrospective version of the BDI was developed and 9-items were selected from this version for inclusion in the RSS.

### New Student Survey

The NSS is composed of 30 self-report items, most of which address students' activities, attitudes, and opinions about the college experience and their health (see Appendix A). Twenty-three of the items require that subjects respond to a 5-point scale on which 1 refers to "strongly disagree" and 5 refers to "strongly agree". Four of the twenty-three items are general health items and were adapted from the RAND 36-Item Health Survey 1.0<sup>3</sup> (Hays, Sherbourne, & Mazel, 1993). Four self-ratings of academic ability are assessed by a 4-point scale on which 1 refers to "not prepared at all" and 4

---

<sup>3</sup> Substantial empirical evidence of validity has accumulated for the four general health items. See also: Ware & Sherbourne, 1992).

refers to “very well prepared”. Two questions regarding health activities require a “yes” or “no” response. One additional opinion variable requires students to rank in order of importance what they believe are the six defining characteristics of an educated person in today’s world.

The items included in the NSS represent five categories of items that have been determined *a priori* to represent five distinct variable types (i.e., factors). The identified variable types are comprised of two opinion variables, two attitudinal variables, and one cognitive variable. One of the opinion variable types focuses on students’ judgments about the college experience, and the other examines students’ beliefs about their academic ability. The two attitudinal variable types identify students’ anxiety about school (i.e., an attitudinal state) and their optimism towards future activities (i.e., an attitudinal trait). The cognitive variable type consists of students’ perceptions of their general health.

The NSS contains items that have been included on surveys previously administered to undergraduates attending a large state university and these items demonstrate acceptable face and content validity. However, to establish the reliability and validity of this instrument, the internal consistency of the items was examined on an independent sample.

### Method

The New Student Survey (NSS) was administered to over 500 students participating in an introductory psychology class at the university. The survey was one of several, self-administered questionnaires that first- and second-year college students are required to complete during their first class meeting to fulfill a course requirement. All

surveys administered in this study were approved by the university Human Subjects Committee, and informed consent was obtained from all participants. From the initial sample, an independent sample of 230 surveys were randomly selected for validation purposes. Although class standing was documented, typical demographic information (e.g., gender and age) was not gathered because characteristics of subjects was not relevant to the methodological focus of this study.

### Data Analysis

The responses to the NSS items (with reversed-keyed items inverted) were grouped into five factors representing the designated variable types<sup>4</sup>. Unit-weighted factor scores (i.e., the mean of the factor items) were constructed and correlated with the individual items comprising the NSS. Correlations between items and factor scores  $> .50$  were to be considered acceptable. The internal consistency of the variable types was determined by using coefficient alpha. Internal consistency estimates were calculated for each of the items in a factor and these estimates denote the contribution that a particular item makes to the total factor.

### Results and Discussion

#### *Factor Confirmation*

Based on the five variable types identified in the NSS, the *a priori* factors constructed were labeled as follows: F1= Judgements about the College Experience

---

<sup>4</sup> Some items in the NSS were not appropriate for inclusion in the construction of variable types (e.g., the rank order of knowledge).

(Experience), F2= Anxiety About School (Anxiety), F3= Optimism Towards Future Activities (Optimism), F4= Beliefs About Academic Ability (Ability), and F5= Perceptions of General Health (Health).

Three of the factors met correlation and internal consistency acceptance criteria: Optimism, Ability, and Health. Although the Experience factor met the correlation criteria, it had poor internal consistency ( $\alpha = .22$ ). The Anxiety factor failed to meet both the correlation and internal consistency inclusion criteria ( $\alpha = -.06$ ). Cases with any missing data were excluded from these analyses. Correlations with the total factor score and the internal consistency estimates for the Optimism factor can be seen in Table 2.

**Table 2: Internal consistency estimates for the Optimism factor.**

Item	N= 225	Correlation with Factor Score	Cronbach Coefficient Alpha= .66
World		.51	
Content		.74	
Social		.69	
Enthusiasm		.63	
Life		.69	

Factor score correlations and the internal consistency estimates for the Ability and Health factors are recorded in Table 3 and Table 4 respectively.

**Table 3: Internal consistency estimates for the Ability factor.**

Item	N= 230	Correlation with Factor Score	Cronbach Coefficient Alpha= .75
Read		.67	
Write		.64	
Organize		.65	
Perform		.72	
Intro class		.60	
GPA		.69	

**Table 4: Internal consistency estimates for the Health factor.**

Item	N= 230	Correlation with Factor Score	Cronbach Coefficient Alpha= .76
Sick		.74	
Healthy		.80	
Worse health		.65	
Excellent health		.85	

*Cross-Validation*

The variable types examined in the initial sample were cross-validated on the Phase 2, NSS study sample containing 270 students. Cases with any missing data were excluded from these analyses. The Experience ( $\alpha = .19$ ) and Anxiety ( $\alpha = .09$ ) factors both failed to meet inclusion criteria, as demonstrated in the previous sample. The remaining three factors, Optimism, Ability, and Health met correlation and internal consistency criteria. Internal consistency estimates for the three factors are reported in Table 5.

**Table 5: Cross-validation of the Optimism, Ability and Health factors.**

Factor	N	Cronbach Coefficient Alpha
Optimism	265	.68
Ability	265	.74
Health	266	.81

**Beck Depression Inventory**

The Beck Depression Inventory (BDI) was selected for inclusion in this study as a measure of “mood state”. The 21-item instrument is regarded as one of the better self-report measures of general depression (Reynolds & Gould, 1981) because it taps the cognitive, affective, motivational, and physiological aspects of depression. The BDI has

been subject to a number of factor analyses over the last two decades. Early studies suggested that items in the BDI tapped three psychologically distinct dimensions (Campbell, Burgess, & Finch, 1984). Welch, Hall, & Walkey, (1990) however, suggest that items on the BDI constitute a general construct of depressive symptoms rather than a three factor solution. Including items from the BDI provides an opportunity to examine memory distortion in an attitudinal variable using a Normed instrument.

In order to reduce respondent burden, 9 of the original 21-items were selected for inclusion in the Retro- and Post- versions of the Student Survey. The 9-items selected for inclusion were items judged by the investigator to be most likely to show the most reasonable variability in a college population.

#### Method

The BDI was administered to 695 students who participated in an introductory psychology class at the university. The survey was one of several, self-administered questionnaires students are required to complete.

#### Data Analysis

The internal consistency for the proposed 9-item scale was determined by coefficient alpha. The internal consistency of the 9-item scale was then compared to the original, 21-item version to determine if the shorter scale was acceptable for inclusion in Phase 2 of the study. The literature indicates that items on the BDI constitute a general construct of depressive symptoms, therefore, the factor structure of the 9-item scale was not checked. Cases with any missing data were excluded from the analysis.

### Results and Discussion

The 9-item BDI scale met internal consistency acceptance criteria (see Table 6).

For the 21-item scale, the coefficient alpha ( $\alpha = .88$ ) and the internal consistency estimates (all in the .87 range) were somewhat higher than those found on the 9-item scale. This was expected because the original instrument contains more items and therefore has higher reliability. The 9-item scale, therefore, is assumed to constitute a general factor of depressive symptoms.

Table 6: Internal consistency estimates for the 9-item BDI scale.

Item	N=615	Correlation with Factor Score	Cronbach Coefficient Alpha= .74
1 Sad		.65	
2 Future		.55	
10 Cry		.60	
11 Irritate		.34	
15 Drive		.63	
16 Sleep		.63	
17 Tired		.63	
18 Appetite		.52	
20 Worried		.48	

#### *Cross-Validation*

The internal consistency of the BDI 9-item scale was cross-validated on the Phase 2, BDI study sample of 235 students. Of the 195 students who reported their class standing, over 90% were first- and second- year students.

Results from the cross-validation study were consistent with results reported in the initial study. The sample size was one-third the size of the initial study which resulted in smaller estimates for both the 21- and 9-item scales. As before, the 21-item scale had a

moderately higher coefficient alpha ( $\alpha = .79$ ). The internal consistency estimates for the 21-items ranged from .77 to .79 and estimates for the 9-item scale ranged from .63 to .68. The internal consistency estimates for the 9-item scale are specified in Table 7.

The 9-item BDI scale constitutes a general depression factor and was therefore included in the Retro- and Post- surveys as a measure of “mood”.

Table 7: Cross-validation of the BDI 9-item scale.

Item	N=205	Correlation with Factor Score	Cronbach Coefficient Alpha= .68
1 Sad		.46	
2 Future		.48	
10 Cry		.61	
11 Irritate		.49	
15 Drive		.46	
16 Sleep		.55	
17 Tired		.60	
18 Appetite		.52	
20 Worried		.54	

## CHAPTER 3: PHASE 2: RETROSPECTIVE PRETEST STUDY

### Study Design

Three survey instruments were administered in the course of this study. The NSS, or pretest, was administered at the first meeting of the psychology course students are enrolled in. The retrospective version, the RSS, and the Post Student Survey (PSS) were administered during the respective time periods indicated in Table 8.

Table 8. Data collection plan.

	<b>PRETEST</b>	<b>RETROSPECTIVE PRETEST</b>	<b>POSTTEST</b>
<b>TIME ADMINISTERED</b>	First class of the semester	RP1: 4 weeks RP2: 8 weeks RP3: 12 weeks	PT1: 4 week PT2: 8 weeks PT3: 12 weeks
<b>SURVEY</b>	<i>New Student Survey Beck Depression Inventory</i>	<i>Retro Student Survey</i>	<i>Post Student Survey</i>

### Method

#### Participants

As previously noted, the NSS was one of several, self-administered questionnaires that students are required to complete during their first class meeting to fulfill a course requirement. Of the more than 500 students who completed the NSS, 235 students selectively enrolled in one of three cohorts. These cohorts then convened at four, eight, or twelve week time points. Over 90% of those who enrolled were first- and second-year university students.

#### Procedure

At each time point, students were asked to complete two surveys. The first survey

administered was the RSS, a retrospective version of the items included in the pretest (see Appendix B). Students were instructed to answer the survey questions in light of how they were “at the beginning of the semester”. Students were instructed to complete the RSS, and when they finished and turned-in the survey, they were given the PSS. The two surveys took students about 20 minutes to complete.

Although no formal intervention is administered thereby making a traditional posttest unnecessary, each cohort of students completed the PSS (see Appendix C) as an additional validity check. The posttest version directs students to report on “how they perceive themselves to be now”. In order to minimize the time students spend completing the retro- and posttest surveys, the posttest form of the survey contains a subset of the items included in both the pretest and retrospective versions. The primary reason for truncating the posttest form was previous evidence suggesting that students may be resistant to the idea of taking a test twice, once “for real” and once “for pretend”. The PSS is designed to capture students’ present state and will help to distinguish those items that are susceptible to response-shift bias.

### Data Analysis

Several analyses were conducted in response to the research questions. The goal of the first phase was to determine the stability and internal consistency of the factor difference scores. In this part of the analysis, the four variable types (or factors) investigated were Optimism, Ability, Health, and Mood States. The retro-minus-pre differences scores were created and the coefficient alpha was calculated to determine the stability of factor differences scores across time.

The aim of the second analysis was to determine the direction and magnitude of change across time for the three combinations of difference scores: post-minus-pre scores, post-minus-retro scores and the retro-pre scores. A regression model was tested using Proc Reg in SAS. The factors were sorted by a variable identified as "Time", and Time was regressed on the each set of the factor difference scores. Difference scores were then plotted.

A general linear model (GLM) was applied in the third analysis to the post-minus-pre scores, post-minus-retro scores, and the retro-minus-pre scores to test for the main effect of Time on three of the factors: Optimism, Ability and Mood. The Health factor was excluded from further analyses because it was lacking posttest data. A GLM model was also used to determine if any of the factors had interactions with Time. The effect of time on pre/post change with and without interactions was tested to identify pretest bias.

In the final analysis, a structural equations model (SEM) was performed using the statistical software program EQS 5.1 (Bentler, 1995). In this particular analysis, a multisample mean and covariance structures application of SEM was applied. The multisample analysis analyzes data from all samples simultaneously, and in this case, each of the variable types was treated as a separate sample. Each of the samples is expected to verify that a model reproduces the sample data of each group to within sampling accuracy. The goodness-of-fit  $\chi^2$  test was used to describe the adequacy of the model. Each of the models uses manifest variables (rather than latent constructs), therefore, raw scores were used in the estimation of path coefficients. By incorporating a mean structure into the model, both path coefficients and the means for each of the variables are estimated. The

means are estimated by including the constant 1 (V999 in EQS) as an independent variable that has no variance and no covariances with other variables in the model (Bentler, 1995, p.167).

Prior to conducting the multisample analysis, the normality of the data was checked using the maximum likelihood robust estimation procedure. Following the multisample analysis, a second multisample model was tested to rule out bias in group selection.

The overall fit of tested models was evaluated using both absolute fit criteria ( $\chi^2$ ) and practical fit of each model (nonnormed fit index (NNFI) (Bentler & Bonnet, 1980) and the robust comparative fit index (CFI) (Bentler, 1989). Models with nonsignificant  $\chi^2$  and/or practical fit indices of .90 or greater were considered acceptable models.

### Results and Discussion

#### Difference Scores

Retro-minus-pre difference scores were created for each of the four variable types. These differences reflect the groups pattern of change in their retrospective reports. Coefficient alpha was calculated for the difference scores to examine how closely the difference scores related to the original alphas (i.e., those calculated on the NSS) and to determine their stability across time. The first cohort contained the most students (N=118) because more students self-selected into this time cohort. The second and third cohorts were comprised of about 74 students each. There were some variations in the number of paired differences scores for each time point because any pair of observations with missing data were removed from the analysis. Table 9 reports the coefficient alphas

for each of the variable types at each of the three time points.

The data indicate that the internal consistency estimates of the Retro/Pre difference scores for the Ability and Health factor were moderately large and consistent across time. The estimates for the Optimism factor became smaller over time. This diminution can be explained, in part, by the fact that the difference scores had less variability (i.e., small differences between retro- and pretest reports on Optimism), so there was less internal consistency to be measured. The Mood factor appears “moody”, that is, the alphas vary in an unsystematic manner across time.

Table 9: Coefficient alpha for the four variable types over time.

Variable Type	Coefficient Alpha		
	T1	T2	T3
Optimism	.44	.31	.08
Ability	.53	.52	.57
Health	.61	.70	.72
Mood	.32	.55	.47

The magnitude and direction of change was determined for the Optimism, Ability, and Mood factors. The Health factor had to be eliminated from further analyses because posttest data were not available. However, the retro-minus-pre difference scores for the Health factor indicated that there were no significant differences between students' self-reports of general health at the 4, 8, or 12 week time points. The retrospective pretest reports of the Health items are not biased by memory distortion at this group level of analysis (i.e., means). Furthermore, since the Health variable type is based on health “traits” which are presumably fairly stable across time, it is not surprising that the retrospective pretest reports appear to be quite similar to pretest reports.

A regression model was used to determine the magnitude and direction of change in the Optimism, Ability, and Mood factors. The means were plotted for each of the three combinations of difference scores, and t-tests identified any significant differences in the paired scores across time. Results for each set of difference scores are displayed in Tables 10, 11, and 12.

**Table 10: Post-minus-pre difference scores across time.**

Variable Type	Means		
	T1	T2	T3
Optimism	.09	.03	.03
Ability	-.09	-.20*	-.18*
Mood	-.19	-.03	-.01

\* Indicates a significant difference between post-minus-pre at the specified time point with  $p \leq .05$ .

Significant changes in the post-minus-pre difference scores for Ability confirms the hypothesis that self-ratings of personal attributes (e.g., academic ability) are more susceptible to subject bias. Since the students received no formal intervention, changes between the posttest and pretest scores can be attributed to subject bias.

**Table 11: Post-minus-retro difference scores across time.**

Variable Type	Means		
	T1	T2	T3
Optimism	.05	.12	.004
Ability	-.10*	-.17*	-.09
Mood	-.07	-.07	-.03

\* Indicates a significant difference between post-minus-retro at the specified time point with  $p \leq .05$ .

The Ability factor exhibits significant, negative differences at the fourth and eighth week for the post-minus-retro scores. The negative sign is indicative of a response-shift. Students rated their academic abilities higher at the time of the pretest, and over the

course of the semester they apparently came to believe that their abilities were not as good as they had originally thought. Therefore, the retrospective reports were altered somewhat. Since the differences scores are based on the means of the groups, it is difficult to identify the exact pattern of response-shift that has occurred, but it is small on average, in any case.

The three factors showed no significant differences in the retro-minus-pre scores across time. In other words, the retrospective scores and the pretest scores were very similar, based on each group's pattern of change.

**Table 12: Retro-minus-pre difference scores across time.**

Variable Type	Means		
	T1	T2	T3
Optimism	-.04	-.07	.02
Ability	-.02	-.03	-.08
Mood	-.04	-.06	.03

\* Indicates a significant difference between retro-minus-pre at the specified time point with  $p \leq .05$  or less.

### The Effect of Time

Results from the GLM indicated there was no main effect of time on any of the pairs of difference scores for the three factors. In addition, there were no interactions found between the three factors and Time for the post-minus-retro scores or the retro-minus-pre scores. One small interaction ( $b=.22$ ) was found in the post-minus-pre scores for the Mood factor and Time. The effect was small enough to be attributable to random error.

There are several plausible explanations for finding no differences across time in the three factors. First, since there was no explicit intervention, there was probably little

change thereby reducing the amount of error that could be introduced into the recall process. Second, both the Optimism and Ability factors are measuring trait-like characteristics that are less susceptible to change, so it is less likely that there should be a need for memory reconstruction. Finally, even though the Mood factor was measuring a state, the degree of depression reported by college students is small and varies little across time intervals.

Given the findings of no differences across time, data from the three time points were collapsed for each factor, which made it possible to combine the data for each factor and treat it as a distinct sample population representing the variable type. The raw data for each of the three samples was then imported into EQS.

#### Structural Equations Models

A non-restricted, saturated model was constructed to identify memory bias in retrospective pretests. Prior to rescaling the instruments to put them in a common metric for the multisample comparison, the model for each of the three samples was subjected to a robust analysis to check for the normality of data. Results indicated that none of the models was sensitive to violations of normality. Robust statistics (e.g., skewness and kurtosis) were nearly identical to the standard errors.

The basic model examines the effect of the pretest on both the posttest and retrospective pretest while taking into account the effect of time for a specific variable type. Means and path coefficients were estimated and are indicated in the model illustrations.

A multisample mean and covariance analysis was conducted to compare the three

models using equality constraints. In order to check for the possibility of bias in group selections, the nonrestricted, saturated model was tested, allowing the error terms for the Pretest and Time to covary. In other words, the model was set as close to "0" as possible (i.e.,  $df=0$ ) in order to illustrate that pretest subjects were randomly selected. Univariate test statistics showed that all constraints were correctly imposed and they had nonsignificant chi-squares. Statistical and practical fit indices reported in Table 13 demonstrate that the model fit the data very well. The model explains the observed correlations to within an average error of .01 based on the average of the absolute values of the standardized residuals.

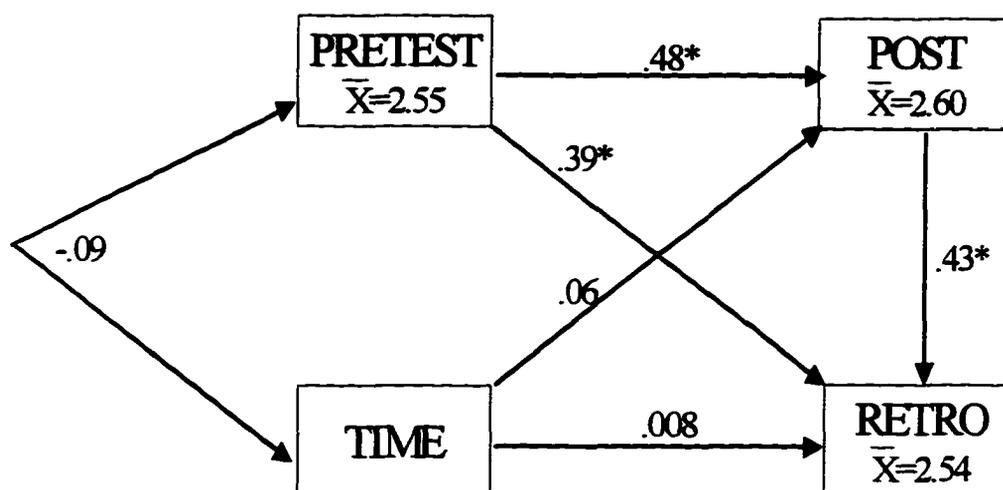
There was not a significant covariance between the error terms for Pretest and Time in any of the models. The results provide evidence that there was no bias present in group assignment for the Optimism, Ability or Mood populations.

**Table 13. Statistical and practical fit indices for the multisample analysis.**

$\chi^2$	df	<i>p</i>	NFI	NNFI	CFI
11.93	12	.45	.98	1.00	1.00

The first model fit was for the Optimism data (see Figure 2). The means reflect group patterns of change; therefore, little difference is noticed between the means for the pretest, posttest, and retrospective pretest. The individual pattern of change represented by the path coefficients provides a different story. For the Optimism variable type, pretest and posttest scores remain fairly stable ( $b=.48$ ). The retrospective pretest scores, however, are not highly reliable when it comes to reproducing the pretest scores for Optimism ( $b=.39$ ). Based on the students' state at posttest, the retrospective pretest

scores indicate that there has been fair amount of memory reconstruction that has occurred ( $b=.43$ ). The different time intervals had no significant effects on either the posttest or the retrospective pretest reports. As mentioned previously, the covariance between pretest and time was not significant.

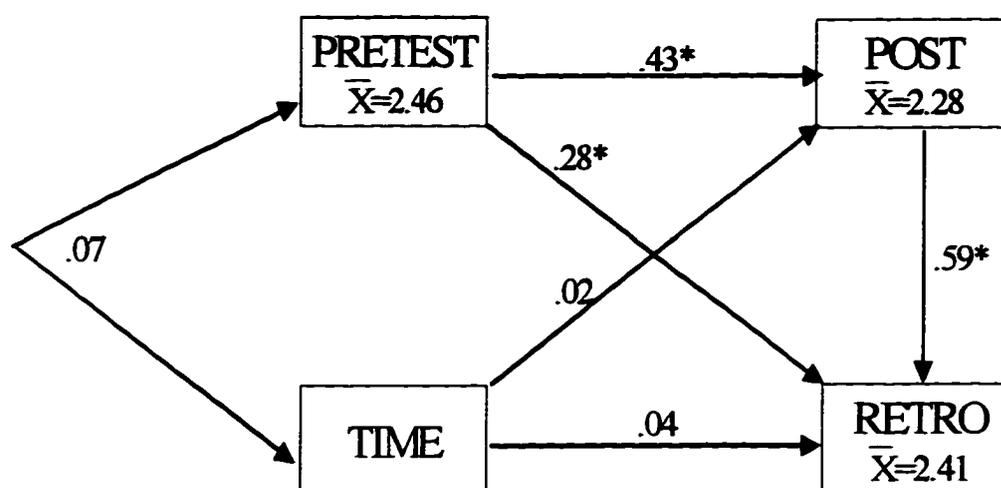


\* $p < .05$

Figure 2: Analytical model of the Optimism variable type.

The second model fit was for the Ability data (see Figure 3). The pretest mean was somewhat larger than both the retrospective pretest and Posttest means. The path between pretest and posttest indicates that scores are stable ( $b=.43$ ). The test/retest reliability represented by the path from pretest to retrospective pretest shows that the two sets of scores are not very reliable ( $b=.28$ ). This lack of reliability is further demonstrated by the path from the posttest to retrospective pretest which exhibits a high degree of memory reconstruction ( $b=.59$ ). As before, there is no significant effect of the time or group bias. A response-shift was hypothesized for this variable type because it was

possible that the school experience itself (i.e., maturation) may serve as a type of intervention thereby changing students' evaluations of their academic abilities. This appears to be the case.



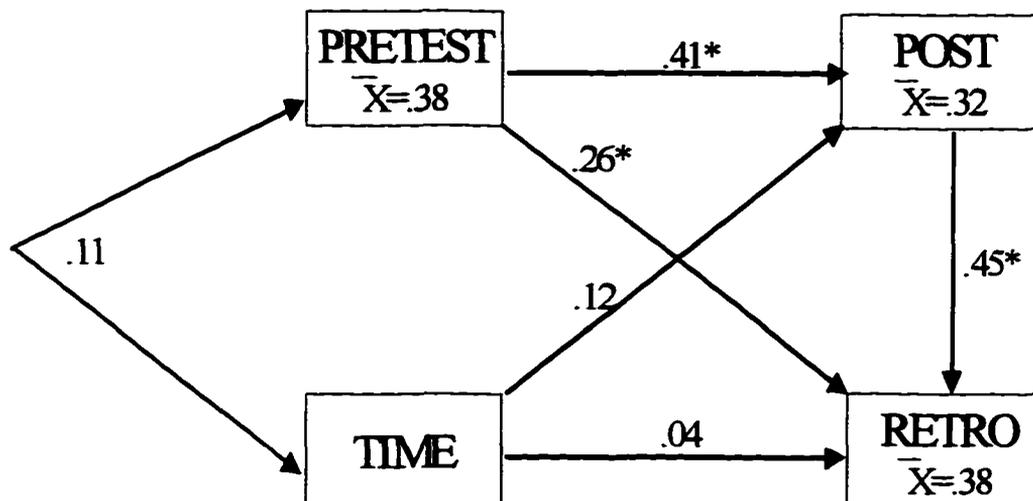
\* $p < .05$

Figure 3: Analytical model of the Ability variable type.

The model representing the Mood factor reports small, identical means for the pretest and retrospective pretest, but the reliability for the two pretests is low ( $b = .26$ ). The Mood scores are about as stable from pretest to posttest as the previous two factors were ( $b = .41$ ). The path coefficient from posttest to the retrospective pretest demonstrates, as it did in the previous two models, that the retrospective pretest scores are being influenced by the present state (i.e., the state students are in when they completed the posttest). A moderate amount of memory reconstruction ( $b = .45$ ) occurs in the minds of students as they retrospectively report on their mood states. There are no effects of time and group bias on the Mood variable type.

The poor test/retest reliability seen in this model is somewhat puzzling when the

test/retest reliability of the BDI items included in this variable type have previously been demonstrated. This finding might be attributable to the small amount of variability that was in the Mood factor and is reflected in the identical means for the pretest and retrospective pretest. The sample population of college students, as noted previously, is less likely to report depressive symptoms. The Mood factor therefore consisted of small differences that may have inadvertently influenced the poor reliability and minimal stability found in this model.



\* $p < .05$

Figure 4: Analytical model of the Mood variable type.

### Individual Items

The NSS also contained single items that were intended to test for the following: hindsight bias, the accuracy of reporting the number of hours spent studying, and the stability of students' opinion about a social issue (e.g., regulating the ownership of guns). Since none of these items were included in variable types previously tested, a separate

analysis was done for these items.

Two questions were asked to test for hindsight bias. First, students were asked to report if they thought that the U of A football team was going to have a really good season (see Appendix A, item #9). Second, students were asked if they thought that U.S. troops would be fighting in Bosnia within the next three months (see Appendix A, item #20). At the time when this pretest question was asked, war in Bosnia appeared to be imminently possible. If hindsight bias were operational, it would most likely be seen in the reports of the last cohort of students. This group would have had the longest opportunity to observe the success or failure of the football team and to observe whether troops had been deployed to Bosnia. For hindsight bias to be operational, the retrospective-minus-pretest difference scores should be significantly different.

The scores for the 3 cohorts are reported in Table 14. Results for the item predicting the success of the football team show that only the second cohort lowered their scores on their retrospective reports. The reason hindsight bias was operational in this cohort rather than the third cohort can most likely be explained by the slump that the football team had in the middle of the season. An additional reason for finding few differences in pretest and retrospective pretest scores for these items can be attributed to the benign position that students took. Most of the means for these items center around the number three on the Likert scale, which stands for "not sure". Students obviously were uninterested in or lacked knowledge about these topics and therefore played it safe by selecting option number three.

**Table 14: Testing for hindsight bias.**

<b>RETRO-MINUS-PRE DIFFERENCE SCORES</b>			
<b>ITEMS</b>	<b>T1</b>	<b>T2</b>	<b>T3</b>
<i>Football Team</i>	-0.11 NS	-0.21 *(.04)	-0.12 NS
<i>War with Bosnia</i>	-0.14 *(.02)	-0.06 NS	-0.02 NS

\* Indicates significance ( $p < .05$ ) for the pair of Difference Scores

The next group of items were examined for specific accuracy. The first item in this group asked students to report the number of hours that they had spent studying during a typical week during their senior year in high school or last year at another college (see Appendix A, item #6). The response set ranged from "1" or None to "5" or more than 10. This item exemplifies the recall of a specific activity, but to answer the question requires the student to use some form of mental aggregation. This particular question offered an opportunity for the investigator to examine whether retrospective responses based on aggregated memory were consistent with pretest reports. The second item asked students to for their opinion on whether regulating the ownership of guns is a threat to our society (see Appendix A, item #17). This item provides an opportunity to examine the accuracy of a retrospective report of a simple opinion. Results for the items are displayed in Table 15.

**Table 15: Accuracy of retrospective pretest reports for specific items.**

<b>RETRO-MINUS-PRE DIFFERENCE SCORES</b>			
<b>ITEMS</b>	<b>T1</b>	<b>T2</b>	<b>T3</b>
<i>Studying</i>	.008	-.06	.08
<i>Regulating Guns</i>	-0.04	.10	-0.14

\* Indicates significance ( $p < .05$ ) for the pair of Difference Scores

The were no significant differences in the retrospective reports for studying or regulating the ownership of guns. Two possible explanations for these results. First, the no difference finding could be attributed to, as was seen in the variable types, the use of group mean difference scores. Second, it is possible that there are no differences in the retrospective reports of these items.

### General Discussion

Results from the analysis of the variable types are contradictory. First, the group difference scores (e.g., retro-minus-pre) for the three cohorts of students indicated that there were no significant differences found across the three time intervals for any of the variable types. The structural equations models that revealed the individual patterns of change for each of the variable types, however, present a different conclusion. The reliability for the pretest and retrospective pretest scores is low, and a moderate amount of memory distortion was found. These findings demonstrate the methodological weakness of using group means to determine systematic bias in retrospective pretest scores. When the pattern of change is examined at the group level, the individual differences are minimized resulting in somewhat erroneous conclusions. For instance, the pretest and retrospective pretest plots for the individual items revealed that, for most items, about half of the students selected the same score on their retrospective pretest as they had indicated on their pretest. The other half of the population was split: students' reports increased or decreased by one or two points on the Likert scale. At any rate, the increases in one or two levels can easily be washed-out by those students whose scores decreased in the same manner. The group difference scores therefore appear to represent the 50% of the

**students who had retrospective reports that were consistent with their pretest reports.**

**The individual patterns of change in the models represent the other half of the population who experienced some degree of memory reconstruction.**

## CHAPTER 4: CONCLUSIONS

The major findings of this study were that there was no main effect of time on any of the pairs of difference scores, and a moderate level of memory reconstruction occurred for the three variable types examined. Based on these findings, several conclusions can be made about the presence of memory distortion in retrospective pretests, the methodology used to detect it, and the limited examination of individual differences.

The surprising lack of significant difference found across the time intervals, may be explained by two plausible explanations. First, since there was no explicit intervention, there was probably nothing that happened across the lives of the students that caused a detectable change to occur. Any change in the lives of students occurred as a result of the daily “college experience”, and it probably did not cause any uniform change. Further research of this type is needed to confirm whether little change results in reducing the amount of error that could be introduced into the recall process (e.g., measuring trait-like characteristics). Second, although twelve weeks may be a sufficient length of time for change to occur, the difference between each of the time points (i.e., four weeks) made it difficult to detect any subtle changes that may have occurred in students’ lives. Further research testing of this methodology is required to determine its usefulness in detecting memory reconstruction across time.

Although a moderate level of memory reconstruction was noted, about 50% of the retrospective self-reports in this study revealed that both attitudinal states and traits as well as opinions are reliable (i.e., same as the pretest reports). The percent of memory reconstruction present in the psychological-oriented variable types does not appear to be

any higher than the percent reported for specific health events. These findings are unlikely to reduce concern over the presence of memory distortion in the retrospective pretest approach. The “good news” to be gleaned from this finding; however, is that retrospective pretests do present a viable alternative for use when pretest data is not available and when a high degree of accuracy is not required. Thus, retrospective pretest data are better than no data, and they are probably reasonably useful with large samples.

The methodology applied in this study offers a productive strategy for identifying instrumentation effects in retrospective pretests at several levels.

First, the current methodology revealed a discrepancy in the traditional approach used to analyze retrospective pretest data. Previous studies have employed the traditional approach, that is, using group means to identify change in post-minus-pre or post-minus-retro scores. The flaw in this approach is that in using group means, individual differences are obscured thereby masking the effects of memory distortion. By examining differences at the individual level through structural equations modeling, the memory distortion associated with the recall of diverse variable types was identified.

Second, the methodology proved useful for testing the magnitude and direction of change in a response shift. The Ability factor contained items that required self-ratings of academic ability, but it is unlikely that these reports were biased since there was no explicit intervention, and students had nothing to gain from inflating their abilities. On the other hand, the Ability factor was the only variable type in this study that was identified by the investigator as having the potential for a response shift. Over the course of 12 weeks, it was possible for students to reassess and gain a new perspective about their academic

skills. Results from both the group and individual patterns of differences verified that more memory reconstruction occurred in the Ability factor. Furthermore, this finding illustrates an additional advantage associated with using the methodology employed in this study. By examining the individual pattern of change, an investigator can test a theory about response-shift or other known biases that may affect the retrospective pretest scores. This approach would provide more support for using post-minus-retrospective scores when a response shift is present than the current practice of selecting the set of difference scores with the largest mean difference.

Third, the analytical models revealed the pattern of relationship between pre-, post-, and retrospective pretests and the significant level of memory reconstruction that occurred in the retrospective pretest for the three variable types.

Future applications of this methodology would benefit by including a heterogeneous sample and investigating a range of complex variables. In addition, investigators need to establish testable theories about the role of memory distortion for specific variable types. These theories should then be tested using normed instruments and previously validated factors. Finally, a better system of classification is needed for reporting bias attributable to memory distortion.

One limitation of this study is that the individual items were not examined for individual subject differences. Future studies of this sort would benefit by identifying the subjects whose retrospective reports change by one or two levels. The responses of these subjects could then be tracked to find out whether the subjects change their responses in a random or systematic manner (e.g., as would result in a change of metric). A better

**understanding of the role of memory distortion in retrospective pretests will enhance our ability to use them under dependable sets of conditions.**

## APPENDIX A: NEW STUDENT SURVEY

## NEW STUDENT SURVEY

NAME \_\_\_\_\_ STUDENT ID# \_\_\_\_\_  
 PHONE \_\_\_\_\_ CLASS STANDING \_\_\_\_\_

How well prepared for college work do you think you are in each of the following areas?

	Not Prepared At All	Not Very Well Prepared	Prepared	Very Well Prepared
1. Ability to read and comprehend psychology texts.	1	2	3	4
2. Ability to write effectively.	1	2	3	4
3. Ability to organize notes and prepare for exams (study skills).	1	2	3	4
4. Ability to perform in class (take notes, take tests).	1	2	3	4

5. Please rank in order of importance what you believe to be the defining characteristics of an educated person in today's world. (Use 1 to indicate the most important and 6 to indicate the least important.)

- General knowledge of the great works of literature, philosophy, and of music and the arts  
 Expertise in at least one discipline (e.g., physics, sociology, history, French, etc.)  
 Practical knowledge of business and political economy  
 Understanding people  
 General knowledge of mathematics, computing, and the sciences  
 Knowledge of more than one language and culture

6. In the year BEFORE you started college at the U of A (senior year of high school or last year at another college for most students), how much time did you spend during a typical week studying/homework?

None     Less than one hour     1-5 hours     6-10 hours     Over 10 hours

Please indicate your opinion on the following statements.

	Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
7. I am highly interested in the field of psychology.	1	2	3	4	5
8. Selecting a major has been difficult for me.	1	2	3	4	5
9. I think that the U of A football team is going to have a really good season.	1	2	3	4	5

	Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
10. I am impressed with the appearance of the U of A campus.	1	2	3	4	5
11. I may have trouble making friends at the U of A.	1	2	3	4	5
12. I am optimistic about the way things are going in the world today.	1	2	3	4	5
13. I think I will do well in the introductory psychology class.	1	2	3	4	5
14. I feel anxious about starting school this semester.	1	2	3	4	5
15. Most of the time, I feel content with my life.	1	2	3	4	5
16. I believe that my grade point average for this semester will be a 3.0 or better.	1	2	3	4	5
17. Regulating the ownership of guns is a threat to our society.	1	2	3	4	5
18. I will have no trouble at all fitting in socially at the U of A.	1	2	3	4	5
19. I consider myself to be an enthusiastic and motivated student.	1	2	3	4	5
20. I believe that U.S. troops will be fighting in Bosnia within the next three months.	1	2	3	4	5
21. I have felt welcomed by faculty at the U of A.	1	2	3	4	5
22. I would say, in general, things in my life are going well.	1	2	3	4	5
23. I may have difficulty keeping up with my school work.	1	2	3	4	5
24. When I enrolled at the U of A, I was concerned about not getting the classes I need to graduate on time.	1	2	3	4	5
<hr/>					
25. I have been to see a dentist in the last 6 months (between March 1 - August 30, 1995).	Yes _____		No _____		
26. During the past month, have you cut down on the amount of time you spent on work or other activities as a result of any emotional problems (such as feeling depressed or anxious)?	Yes _____		No _____		
<hr/>					

Please choose the answer that best describes how true or false each of the following statements is for you.

	<b>Definitely False</b>	<b>Mostly False</b>	<b>Not Sure</b>	<b>Mostly True</b>	<b>Definitely True</b>
27. I seem to get sick a little easier than other people.	1	2	3	4	5
28. I am as healthy as anybody I know.	1	2	3	4	5
29. I expect my health to get worse.	1	2	3	4	5
30. My health is excellent.	1	2	3	4	5

---

## APPENDIX B: BECK DEPRESSION INVENTORY (BDI)

NAME \_\_\_\_\_ PHONE # \_\_\_\_\_

**On this questionnaire are groups of statements. Please read each group of statements carefully. Then pick out the one statement in each group which best describes the way you have been feeling the past week, including Today. Circle the number beside the statement you picked. If several statement in the group seem to apply equally well, circle each one. Be sure to read all the statements in each group before making your choice.**

1.     0 I do not feel sad.  
       1 I feel sad.  
       2 I am sad all the time and I can't snap out of it.  
       3 I am so sad or unhappy that I can't stand it.
  
2.     0 I am not particularly discouraged about the future.  
       1 I felt discouraged about the future.  
       2 I felt I had nothing to look forward to.  
       3 I felt that the future is hopeless and that things could not improve.
  
3.     0 I do not feel like a failure.  
       1 I feel I have failed more than the average person.  
       2 As I look back on my life, all I can see is a lot of failures.  
       3 I feel I am a complete failure as a person.
  
4.     0 I get as much satisfaction out of things as I used to.  
       1 I don't enjoy things the way I used to.  
       2 I don't get real satisfaction out of anything anymore.  
       3 I am dissatisfied or bored with everything.
  
5.     0 I don't feel particularly guilty.  
       1 I feel guilty a good part of the time.  
       2 I feel quite guilty most of the time.  
       3 I feel guilty all of the time.
  
6.     0 I don't feel I am being punished.  
       1 I feel I may be punished.  
       2 I expect to be punished.  
       3 I feel I am being punished.
  
7.     0 I don't feel disappointed in myself.  
       1 I am disappointed in myself.  
       2 I am disgusted with myself.  
       3 I hate myself.
  
8.     0 I don't feel I am any worse than anybody else.  
       1 I am critical of myself for my weaknesses or mistakes.  
       2 I blame myself all the time for my faults.  
       3 I blame myself for everything bad that happens.

9. 0 I don't have any thoughts of killing myself.  
1 I have thoughts of killing myself.  
2 I would like to kill myself.  
3 I would kill myself if I had the chance.
10. 0 I don't cry any more than usual.  
1 I cry now more than I used to.  
2 I cry all the time now.  
3 I used to be able to cry, but now I can't even though I want to.
11. 0 I am no more irritated now than I ever was.  
1 I get annoyed or irritated more easily than I used to.  
2 I feel irritated all the time.  
3 I don't get irritated at all by the things that used to irritate me.
12. 0 I have not lost interest in other people.  
1 I am less interested in other people than I used to be.  
2 I have lost most of my interest in other people.  
3 I have lost all of my interest in other people.
13. 0 I make decisions about as well as I ever could.  
1 I put off making decisions more than I used to.  
2 I have greater difficulty in making decisions than before.  
3 I can't make decisions at all anymore.
14. 0 I don't feel I look any worse than I used to.  
1 I am worried I am looking old or unattractive.  
2 I feel that there are permanent changes in my appearance that make me look unattractive.  
3 I believe the I look ugly.
15. 0 I can work about as well as before.  
1 It takes an extra effort to get started at doing something.  
2 I have to push myself very hard to do anything.  
3 I can't do any work at all.
16. 0 I can sleep as well as usual.  
1 I don't sleep as well as I used to.  
2 I wake up 1-2 hours earlier than usual and find it hard to get back to sleep.  
3 I wake up several hours earlier than I used to and can not get back to sleep.
17. 0 I don't get more tired than usual.  
1 I get tired more easily than I used to.  
2 I get tired from doing almost anything.  
3 I am too tired to do anything.
18. 0 My appetite is no worse than usual.  
1 My appetite is not as good as it used to be.  
2 My appetite is much worse then.  
3 I had no appetite at all then.

19. 0 I haven't lost much weight, if any, lately.  
1 I have lost more than 5 pounds.  
2 I have lost more than 10 pounds.  
3 I have lost more than 15 pounds.  
I am purposely trying to lose weight by eating less. YES \_\_\_\_\_ NO \_\_\_\_\_
20. 0 I am no more worried about my health than usual.  
1 I am worried about physical problems such as aches and pains; or upset stomach; or constipation.  
2 I am very worried about physical problems and it is hard to think of much else.  
3 I am so worried about my physical problems that I could not think about anything else.

## APPENDIX C: RETROSPECTIVE STUDENT SURVEY (RSS)

### RETRO STUDENT SURVEY "How I Was at When the Semester Began"

NAME----- STUDENT ID #-----

**Psychologists are interested in many aspects of memory. This questionnaire has to do with your recollection of things you have experienced and the way you have thought and felt in the recent past. Your answers will be held in the strictest confidence and will only be used for statistical analyses. It will take approximately 15 minutes to complete this form. By completing this survey, you are indicating your consent to participated in this study. Thank you for your time.**

As you now see it, how well prepared for college work were you in each of the following areas?

	Not Prepared At All	Not Very Well Prepared	Prepared	Very Well Prepared
1. Ability to read and comprehend psychology texts.	1	2	3	4
2. Ability to write effectively.	1	2	3	4
3. Ability to organize notes and prepare for exams (study skills).	1	2	3	4
4. Ability to perform in class (take notes, take tests).	1	2	3	4

5. Please rank in order of importance what you believed to be the defining characteristics of an educated person in today's world when you first began this semester. (Use 1 to indicate the most important and 6 to indicate the least important.)

- \_\_\_\_\_ General knowledge of the great works of literature, philosophy, and of music and the arts
- \_\_\_\_\_ Expertise in at least one discipline (e.g., physics, sociology, history, French, etc.)
- \_\_\_\_\_ Practical knowledge of business and political economy
- \_\_\_\_\_ Understanding people
- \_\_\_\_\_ General knowledge of mathematics, computing, and the sciences
- \_\_\_\_\_ Knowledge of more than one language and culture

6. In the year BEFORE you started college at the U of A (senior year of high school or last year at another college for most students), how much time did you spend during a typical week studying/homework?

\_\_\_\_\_ None    \_\_\_\_\_ Less than one hour    \_\_\_\_\_ 1-5 hours    \_\_\_\_\_ 6-10 hours    \_\_\_\_\_ Over 10 hours

<b>As you now see it, when Fall semester began:</b>	<b>Strongly Disagree</b>	<b>Disagree</b>	<b>Not Sure</b>	<b>Agree</b>	<b>Strongly Agree</b>
7. I was highly interested in the field of psychology.	1	2	3	4	5
8. Selecting a major had been difficult for me.	1	2	3	4	5
9. I thought that the U of A football team was going to have a really good season.	1	2	3	4	5
10. I was impressed with the appearance of the U of A campus.	1	2	3	4	5
11. I thought I might have trouble making friends at the U of A.	1	2	3	4	5
12. I was optimistic about the way things were going in the world today.	1	2	3	4	5
13. I thought I would do well in the introductory psychology class	1	2	3	4	5
14. I felt anxious about starting school this semester.	1	2	3	4	5
15. Most of the time, I felt content with my life.	1	2	3	4	5
16. I believed that my grade point average for Fall semester would be a 3.0 or better.	1	2	3	4	5
17. I thought regulating the ownership of guns was a threat to our society.	1	2	3	4	5
18. I believed I would have no trouble at all fitting in socially at the U of A.	1	2	3	4	5
19. I considered myself to be an enthusiastic and motivated student.	1	2	3	4	5
20. I believed that United States troops would be fighting in Bosnia within the next three months.	1	2	3	4	5
21. I felt welcomed by faculty at the U of A.	1	2	3	4	5
22. In general, I thought things in my life were going well.	1	2	3	4	5
23. I thought I 'd have difficulty keeping up with my school work.	1	2	3	4	5
24. When I enrolled at the U of A, I was concerned about not getting the classes I need to graduate on time.	1	2	3	4	5

25. I went to the dentist in the 6 month time period between March 1 - August 30, 1995. Yes \_\_\_\_\_ No \_\_\_\_\_

26. When Fall semester began, I had cut down on the amount of time I spent on work or other activities as a result of any emotional problems (such as feeling depressed or anxious)? Yes \_\_\_\_\_ No \_\_\_\_\_

---

Please choose the answer that best describes how true or false each of the following statements was for you when **Fall semester began**.

	Definitely False	Mostly False	Not Sure	Mostly True	Definitely True
27. I seem to get sick a little easier than other people.	1	2	3	4	5
28. I am as healthy as anybody I know.	1	2	3	4	5
29. I expect my health to get worse.	1	2	3	4	5
30. My health is excellent.	1	2	3	4	5

---

Please read the following groups of statements carefully. Then pick out the one statement in each group that best describes the way you felt when **Fall Semester began this past August**. Circle the number beside the statement you picked. If several statements in the group seem to apply equally well, circle each one. Be sure to read all the statements in each group before making your choice.

31. 0 I did not feel sad.  
1 I felt sad.  
2 I was sad all the time and I couldn't snap out of it.  
3 I was so sad or unhappy that I couldn't stand it.
32. 0 I was not particularly discouraged about the future.  
1 I felt discouraged about the future.  
2 I felt I had nothing to look forward to.  
3 I felt that the future was hopeless and that things could not improve.
33. 0 I got as much satisfaction out of things as I used to.  
1 I didn't enjoy things the way I used to.  
2 I didn't get real satisfaction out of anything anymore.  
3 I was dissatisfied or bored with everything.
34. 0 I didn't feel disappointed in myself.  
1 I was disappointed in myself.  
2 I was disgusted with myself.  
3 I hated myself.
35. 0 I didn't cry any more than usual.  
1 I cried then more than I used to.  
2 I cried all the time then.  
3 I used to be able to cry, but then I couldn't even though I wanted to.

36. 0 I was no more irritated then than I ever was.  
 1 I got annoyed or irritated more easily than I used to.  
 2 I felt irritated all the time.  
 3 I didn't get irritated at all by the things that used to irritate me.
37. 0 I had not lost interest in other people.  
 1 I was less interested in other people than I used to be.  
 2 I had lost most of my interest in other people.  
 3 I had lost all of my interest in other people.
38. 0 I couldn't work as well as before.  
 1 It took an extra effort to get started at doing something.  
 2 I had to push myself very hard to do anything.  
 3 I couldn't do any work at all.
39. 0 I could sleep as well as usual.  
 1 I didn't sleep as well as I used to.  
 2 I woke up 1-2 hours earlier than usual and found it hard to get back to sleep.  
 3 I woke up several hours earlier than I used to and could not get back to sleep.
40. 0 I wasn't more tired than usual.  
 1 I got tired more easily than I used to.  
 2 I got tired from doing almost anything.  
 3 I was too tired to do anything.
41. 0 My appetite was no worse than usual.  
 1 My appetite was not as good as it used to be.  
 2 My appetite was much worse then.  
 3 I had no appetite at all then.
42. 0 I was no more worried about my health than usual.  
 1 I was worried about physical problems such as aches and pains; or upset stomach; or constipation.  
 2 I was very worried about physical problems and it was hard to think of much else.  
 3 I was so worried about my physical problems that I could not think about anything else.

When Fall semester began....

43. had you declared a major yet? \_\_\_\_\_ YES \_\_\_\_\_ NO

44. how committed were you to this major?

1 2 3 4 5 6 7 8 9  
 not at all committed moderately committed totally committed

## APPENDIX D: POST STUDENT SURVEY (PSS)

**POST STUDENT SURVEY**  
*"How I Am Today"*

NAME----- STUDENT ID #-----

The items on the following two pages concern your present condition or how you feel today. Your response to these items may be different from previous responses you have given, and this is OK. Many people have views that change over time. Please read the instructions before each set of statements and respond accordingly.

Please rate your ability to complete college work in each of the following areas.

	Very Poor			Excellent
1. Ability to read and comprehend psychology texts.	1	2	3	4
2. Ability to write effectively.	1	2	3	4
3. Ability to organize notes and prepare for exams (study skills).	1	2	3	4
4. Ability to perform in class (take notes, take tests).	1	2	3	4

Indicate your opinion on the following statements to show how you feel right now.

	Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
5. I am highly interested in the field of psychology.	1	2	3	4	5
6. I am impressed with the appearance of the U of A campus.	1	2	3	4	5
7. I have trouble making friends at the U of A.	1	2	3	4	5
8. I am optimistic about the way things are going in the world today.	1	2	3	4	5
9. I am doing well in the introductory psychology class.	1	2	3	4	5
10. Most of the time, I am content with my life.	1	2	3	4	5
11. I believe that my grade point average for this semester will be a 3.0 or better.	1	2	3	4	5

	<b>Strongly Disagree</b>	<b>Disagree</b>	<b>Not Sure</b>	<b>Agree</b>	<b>Strongly Agree</b>
12. I have had no trouble at all fitting in socially at the U of A.	1	2	3	4	5
13. I feel welcomed by faculty at the U of A.	1	2	3	4	5
14. I have had difficulty keeping up with my school work.	1	2	3	4	5

---

**Please read the following groups of statements carefully. Then pick out the one statement in each group that best describes the way you feel today. Circle the number beside the statement you picked. If several statements in the group seem to apply equally well, circle each one. Be sure to read all the statements in each group before making your choice.**

15.    0 I do not feel sad.  
      1 I feel sad.  
      2 I am sad all the time and I can't snap out of it.  
      3 I am so sad or unhappy that I can't stand it.
16.    0 I am not particularly discouraged about the future.  
      1 I feel discouraged about the future.  
      2 I feel I had nothing to look forward to.  
      3 I feel that the future is hopeless and that things cannot improve.
17.    0 I don't cry any more than usual.  
      1 I cry now more than I used to.  
      2 I cry all the time then.  
      3 I used to be able to cry, but now I can't even though I want to.
18.    0 I am no more irritated now than I ever was.  
      1 I get annoyed or irritated more easily than I used to.  
      2 I feel irritated all the time.  
      3 I don't get irritated at all by the things that used to irritate me.
19.    0 I can work as well as before.  
      1 It takes an extra effort to get started at doing something.  
      2 I have to push myself very hard to do anything.  
      3 I can't do any work at all.
20.    0 I can sleep as well as usual.  
      1 I don't sleep as well as I used to.  
      2 I wake up 1-2 hours earlier than usual and find it hard to get back to sleep.  
      3 I wake up several hours earlier than I used to and can not get back to sleep.
21.    0 I don't get more tired than usual.  
      1 I get tired more easily than I used to.  
      2 I get tired from doing almost anything.  
      3 I am too tired to do anything.

22. 0 My appetite is no worse than usual.  
1 My appetite is not as good as it used to be.  
2 My appetite is much worse now.  
3 I have no appetite at all anymore.
23. 0 I am no more worried about my health than usual.  
1 I am worried about physical problems such as aches and pains; or upset stomach; or constipation.  
2 I am very worried about physical problems and it was hard to think of much else.  
3 I am so worried about my physical problems that I cannot think about anything else.
-

## REFERENCES

- Aiken, L.S., & West, S.G. (1990). Invalidation of true experiments: Self-report pretest biases. Evaluation Review, 14, 374-390.
- Armenakis, A.A., Buckley, M.R., & Bedeian, A.G. (1986). Survey research measurement issues in evaluating change: A laboratory investigation. Applied Psychological Measurement, 10, 147-157.
- Bauman, K.E., & Dent, C.W. (1982). Influence of an objective measure on self-reports of behavior. Journal of Applied Psychology, 67, 623-628.
- Bedeian, A.G., Armenakis, A.A., & Gibson, R.W. (1980). The measurement and control of beta change. Academy of Management Review, 5, 1-566.
- Bentler, P.M. (1989-1995). EQS structural equations program manual. Los Angeles: BMDP Statistical Software.
- Bentler, P.M. & Bonnet, D.G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structure. Psychological Bulletin, 88, 588-606.
- Bergen, A.E. (1971). The evaluation of therapeutic outcomes. In A.E. Bergen & S.L. Garfield (Eds.), Handbook of psychotherapy and behavior change. New York: John Wiley & Sons.
- Bradburn, N., Rips, L., & Shevell, S. (1987). Answering autobiographical questions: The impact of memory and inference on surveys, Science, 236, 157-161.
- Buckhout, R. (1974). Eyewitness testimony. Scientific American, 231, 23-31.
- Buckhout, R., Fox, P., & Rabinowitz, M. (1989). Estimating the duration of an earthquake: Some shaky field observations. Bulletin of the Psychonomic Society, 27, 375-378.
- Calsyn, R.J., Allen, G., Morse, G.A., Smith, R. & Templehoff, B. (1993). Can you trust self-report data provided by homeless mentally ill individuals? Evaluation Review, 17, 353-366.
- Campbell, D.T., & Stanley, J.C. (1963). Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally.

Campbell, I. M., Burgess, P.M., & Finch, S.J. (1984). A factorial analysis of BDI scores. Journal of Clinical Psychology, 40(4), 992-996.

Campbell, J.D. & Tesser, A. (1983). Motivational interpretations of hindsight bias: An individual difference analysis. Journal of Personality, 51, 605-620.

Catina, J.A., Gibson, D.R., Chitwood, D.D., & Coates, T.T. (1990). Methodological problems in AIDS behavioral research: Influences on measurement error and participation bias in studies of sexual behavior. Psychological Bulletin, 108, 339-362.

Christensen-Szalanski, J.J.J., & William, C.F. (1991). The hindsight bias: A meta-analysis. Organizational Behavior and Human Decision Processes, 48, 147-168.

Collins, L. M., Graham, J.W., Hansen, W.B., & Johnson, A.C. (1985). Agreement between retrospective accounts of substance use and earlier reported substance use. Applied Psychological Measurement, 9, 301-309.

Conway, M. (1990). On bias in autobiographical recall: Retrospective adjustments following disconfirmed expectations. Journal of Social Psychology, 130, 183-189.

Conway, M., & Ross, M. (1984). Getting what you want by revising what you had. Journal of Personality and Social Psychology, 47, 738-748.

Crockett, L.H., Schulenberg, J.E., & Petersen, A.C. (1987). Congruence between objective and self-report data in a sample of young adolescents. Journal of Adolescent Research, 2, 383-392.

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

Deutsch, M., & Collins, M.E. (1951). Interracial housing: A psychological evaluation of a social experiment. Minneapolis: University of Minnesota Press.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. Journal of Experimental Psychology: Human Perception and Performance, 1, 288-299.

Golembiewski, R. T., Billingsley, K., & Yeager, S. (1976). Measuring change and persistence in human affairs: Types of change generated by O D designs. Journal of Applied Behavioral Science, 12, 133-157.

Hays, R.D., Sherbourne, C. D., & Mazel, R.M. (1993). The RAND 36-item health survey 1.0. Health Economics, 2, 217-227.

Hoogstraten, J. (1982). The retrospective pretest in an educational training context. Journal of Experimental Education, 50, 200-204.

Hoogstraten, J. (1985). Influence on objective measures on self-reports in a retrospective pretest- posttest design. Journal of Experimental Education, 53, 207-210.

Hoogstraten, J., & Koele, P. (1988). A method for analyzing retrospective pretest/posttest designs: II. Application. Bulletin of the Psychonomic Society, 26, 124-125.

Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. Evaluation Review, 4, 93-106.

Howard, G.S., & Dailey, P.R. (1979). Response-shift bias: A source of contamination of self-report measures. Journal of Applied Psychology, 64, 144-150.

Howard, G.S., Dailey, P.R., & Gulanick, N.A. (1979). The feasibility of informed pretests in attenuating response-shift bias. Applied Psychological Measurement, 3, 481-494.

Howard, G.S., Ralph K.M., Gulanick, N.A., Maxwell, S.E., Nance, D.W., & Gerber, S.K. (1979). Internal invalidity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. Applies Psychological Measurement, 3, 1-23.

Howard, G.S., Schmeck, R.R., & Bray, J.H. (1979). Internal invalidity in studies employing self-report instruments: A suggested remedy. Journal of Educational Measurement, 16, 129-135.

Howard, G.S., Maxwell, S.E., Wiener, R.L., Boynton, K.S., & Rooney, W.M. (1980). Is a behavioral measure the best estimate of behavioral parameters? Perhaps not. Applied Psychological Measurement, 4, 293-311.

Howard, G.S., Millham, J., Slaten, S., & O'Donnell, L. (1981). Influence of subject response style effects on retrospective measures. Applied Psychological Measurement, 5, 89-100.

Koele, P., & Hoogstraten, J. (1988). A method for analyzing retrospective pretest/posttest designs: I. Theory. Bulletin of the Psychonomic Society, 26, 51-54.

Kotre, J. (1995). White Gloves. How We Create Ourselves Through Memory. New York: Free Press.

Loftus, E. (1979). Eyewitness testimony. Cambridge, Mass: Harvard University Press.

Mark, M.M. & Mellor, S. (1991). Effect of self-relevance of an event on hindsight bias: The foreseeability of a layoff. Journal of Applied Psychology, 76(4), 560-577.

Nicholson, T., Belcastro, P.A., & Gold, R.S. (1985). Retrospective pretest-posttest analysis versus traditional pretest-posttest analysis. Psychological Reports, 57, 525-526.

Pohl, N.F. (1982). Using retrospective pre-ratings to counteract response-shift confounding. Journal of Experimental Education, 50, 211-214.

Porras, J.I. & Singh, J.V. (1986). Alpha, beta, and gamma change in modeling-based organization development. Journal of Occupational Behavior, 7, 9-24.

Rippey, R.M., Geller, L.M., & King, D.W. (1978). Retrospective pretesting in the cognitive domain. Evaluation Quarterly, 2, 481-491.

Ross, M., McFarland, C., Conway, M., & Zanna, M.P. (1983). Reciprocal relation between attitudes and behavior recall: Committing people to newly formed attitudes. Journal of Personality and Social Psychology, 45, 257-267.

Schmitt, N. (1982). The use and analysis of covariance structures to assess beta and gamma change. Multivariate Behavioral Research, 17, 343-358.

Sears, R.R., MacCoby, E.E., & Levin, H. (1957). Patterns of Child Rearing. Evanston, IL: Row, Peterson.

Southwick, S.M., Morgan, C.A., III, Nicolaou, A.L. & Charney, D.S. (1997). Consistency of memory for combat-related traumatic events in veterans of Operation Desert Storm. American Journal of Psychiatry, 154(2), 173-177.

Sprangers, M. (1988). Response shift and the retrospective pretest: On the usefulness of retrospective pretest-posttest designs in detecting training related shifts. Het Instituut voor Onderzoek van het Onderwijs; Den Haag, Netherlands;116.

Sprangers, M. (1989). Subject bias and the retrospective pretest in retrospect. Bulletin of Psychonomic Society, 27, 11-14.

Sprangers, M., & Hoogstraten, J. (1987). Response-style effects, response-shift bias and a bogus pipeline. Psychological Reports, *61*, 579-585.

Sprangers, M., & Hoogstraten, J. (1988a). Response-style effects, response-shift bias, and a bogus-pipeline: A replication. Psychological Reports, *62*, 11-16.

Sprangers, M., & Hoogstraten, J. (1988b). Subject bias in three self-report measures of change. In Sprangers, Response shift and the retrospective pretest: On the usefulness of retrospective pretest-posttest designs in detecting training related shifts (pp. 57-70).

Sprangers, M., & Hoogstraten, J. (1988c). On delay and reassessment of retrospective preratings. Journal of Experimental Education, *56*, 148-153.

Sprangers, M., & Hoogstraten, J. (1989). Pretesting effects in retrospective pretest-posttest design. Journal of Applied Psychology, *74*, 265-272.

Sprangers, M., & Hoogstraten, J. (1991). Subject bias in three self-report measures of change. Methodika, *5*, 1-13.

Terborg, J.R., & Davis, G.A. (1982). Evaluation of a new method for assessing change to planned job redesign as applied to Hackman and Oldham's job characteristic model. Organizational Behavior & Human Performance, *29*, 112-128.

Terborg, J.R., Maxwell, S.E., & Howard, G.S. (1982). On the measurement and control of beta change: Problems with the Bedeian, Armenakis, and Gibson technique. Academy of Management Review, *7*, 292-295.

Toedter, L.J., Lasker, J.N., & Campbell, D.T. (1990). The comparison group problem in bereavement studies and the retrospective pretest. Evaluation Review, *14*, 75-90.

Ware, J.E., & Sherbourne, C.D. (1992). The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. Medical Care, *30*, 473-483.

Welch, G., Hall, A., & Walkey, F. (1990). The replicable dimensions of the Beck Depression Inventory. Journal of Clinical Psychology, *46*(6), 817-827.

Zwiebel, A. (1987). Changing educational counselors' attitudes toward mental retardation: Comparison of two measurement techniques. International Journal of Rehabilitation Research, *10*, 383-389.