

THESIS



3 1293 02058 6321

2

2000

This is to certify that the

dissertation entitled

INVESTIGATING THE EFFECTS OF ITEM WORDING
ON RATING RESPONSES

presented by

Annie Woo

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Measurement and
Quantitative Methods

William McBrown
Major professor

Date Nov 15, 1999

LIBRARY
Michigan State
University

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

DATE DUE	DATE DUE	DATE DUE

**INVESTIGATING THE EFFECTS OF ITEM WORDING
ON RATING RESPONSES**

By

Annie Woo

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Counseling, Educational Psychology and Special Education

1999

ABSTRACT

INVESTIGATING THE EFFECTS OF ITEM WORDING ON RATING RESPONSES

By

Annie Woo

The purpose of the study was to investigate how item wording affected rating responses. Semantic negative and positive items measuring the same construct, in the form of a Likert-type scale, were tested on a sample of students enrolled in middle schools in Taipei, the capital of Taiwan. The psychometric properties of each item (means, standard deviation, skewness, and kurtosis) were examined as a function of four modes of item wording. The four modes were: Mode 1 (regular), "I like myself"; Mode 2 (negated), "I do not like myself"; Mode 3 (polar opposite), "I dislike myself"; and Mode 4 (negated polar opposite), "I do not dislike myself." A hierarchical measurement model regarding the relationship between the modes of item wording and the responses was constructed. A hierarchical confirmatory analysis estimated the correlations between the item scores and the four modes of item wording. The correlations between four versions of a shame scale and a measure of anxiety (DOSC Anxiety Factor) and of life satisfaction (Satisfaction with Life Scale) were also computed.

Pearson correlation coefficients were obtained for each of these subscales (Mode 1 to Mode 4), DOSC-Anxiety Factor, Satisfaction with Life Scale, and gender. The

Pearson correlation coefficients ranged from -0.133 to 0.898. Gender had low correlation coefficients with the four modes, DOSC, and Satisfaction with Life Factor, ranging from -0.133 to 0.128. There seems to be not much relationship between gender and these scales.

The MANOVA results showed that the responses to the modes of item wording are significantly different between male and female. The F-ratios of Mode 1, Mode 2, and Mode 3 were all significant at the 0.05 level. Mode 4 was not significant at the 0.05 level. These results were similar to that of the correlational analyses. It seems that Mode 4, which has double negative semantics, introduced some ambiguity to the items.

To determine whether the four Modes of semantics measured the same construct, five models were tested in a confirmatory factor analysis. The 2-factor model (Modes 1, 2, & 3 vs. Mode 4) fit the data statistically and showed an overwhelming superiority over the other models. These results rendered strong indications of the inequivalence between double negatives (Mode 4) and the rest of the items (Modes 1, 2, & 3).

Though it may be useful to include some negative items to reduce a response bias, the findings from the present study suggest that special caution should be exercised in the use of double negative item phrasing. Despite the conventional wisdom so often found in measurement textbooks, recent findings by researchers in the area of item phrasing have suggested that negatively phrased items, especially double negatives, may reduce the validity of a questionnaire. The present study clearly corroborates this position.

Copyright by
ANNIE WOO
1999

Dedicated to

Paul and Nathan

ACKNOWLEDGMENTS

This dissertation has been completed with the help of many people. First of all, I wish to express my gratitude to the members of my Dissertation Committee. I wish to thank Dr. William Mehrens, who served as the dissertation chairmen of my committee, for his concerns and guidance in the completion of this study. His friendship and willingness to offer suggestions are most appreciated. I also wish to thank, Dr. Betsy Becker for her statistical expertise and suggestions in the data analysis, Dr. Teresa Tatto for her assistance during the all stages of my study, and Dr. Margot Kurtz for her constructive suggestions as a member of the committee.

I am ever grateful to Margaret Gunn for her assistance in the editing and proofreading of this dissertation.

I wish to express my appreciation also to my best friend, Christine Lau for her encouragement and assistance at some of the crucial stages of my study. Without her help, this study would have taken a much longer time to complete.

I especially appreciate the support and cooperation of all the students, teachers, and school administrators who participated in this study. I will always remember their contributions of time and effort.

Finally and most of all, I wish to thank my husband, Paul, for his love, patience, understanding, and support during the completion of my dissertation. I appreciate the many sacrifices made by him to make it all possible.

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	ix
LIST OF FIGURES.....	xi

CHAPTER

I. STATEMENT OF THE PROBLEM

Introduction	1
Purpose of the Study.....	4
Significance of the Study.....	4
Research Questions	7
Overview	8

II. REVIEW OF LITERATURE

Introduction	9
Response Style.....	9
Item Wording.....	12
Counterbalance of Positive and Negative Items.....	12
Item Reversal.....	16
Summary.....	24

III. RESEARCH DESIGN AND PROCEDURE

Introduction	29
Development of the Test Instrument	29
General Shame Scale.....	29
Dimensions of Self-Concept (DOSC)	31
Satisfaction with life Scale	32
Questionnaire Development	33
Selection of the Sample	37
Procedures of Test Administration.....	37
Quality Control Screening	38
Data Analyses	38
Confirmatory Factor Analysis	40
Exploratory Factor Analysis	41
Parallelism	42
Summary.....	42

IV. ANALYSIS AND INTERPRETATION OF THE DATA

Introduction	44
Characteristics of the Sample	44
Analyses of the Questionnaire Data	46
Answers to the Research Questions	48
Research Question 1	48
Histogram Analyses	48
Reliability Analyses	52
Correlational Analyses	53
Research Question 2	57
MANOVA Analyses	57
Exploratory Analyses	60
Confirmatory Analyses	63
Summary.....	66

V. SUMMARY, CONCLUSIONS, IMPLICATIONS, AND RECOMMENDATIONS

Summary of the Purpose and Procedures of the Study	67
Discussion and Conclusion	68
Implications	70
Cross Validation Using a New Sample	74
Cultural Specificity of the Questionnaire	76
Issues Concerning the Use of Tests.....	78
Limitations.....	79
Recommendations	80

APPENDICES

A. Four modes of Item Wording.....	83
B. Sample Questionnaire in English.....	86
C. Sample Questionnaire in Chinese	94
D. Descriptive Statistics of the Items.....	102
E. Descriptive Statistics of Four Modes of Item Wording	106
F. Item-Total Statistics	109

BIBLIOGRAPHY	112
--------------------	-----

LIST OF TABLES

	Page
Table 1. Summary of Findings Related to the Study	23-24
Table 2. The General Shame Scale (Chang & Hunter, 1988)	35
Table 3. Dimensions of Self-Concept (DOSC) (Anxiety Factor) (Michael & Smith, 1976).....	36
Table 4. Satisfaction with Life Scale (Diener, et al., 1985).....	36
Table 5. Gender of the Participants	45
Table 6. Age of the Participants	45
Table 7. Grade Level of the Participants	45
Table 8. Participants' Age and Gender by Grade	46
Table 9. Coding Format of the Subscales in the Questionnaire	47
Table 10. Tests of Normality	49
Table 11. Results of Reliability Analysis on Four Modes of General Shame Scale (88 Items) in the Questionnaire	52
Table 12. Mean, Standard Deviation, and Cronbach's Alpha Coefficients for the Subscales.....	53
Table 13. Correlations between the 4 Modes, DOSC-Anxiety Factor, Satisfaction with Life Scale, and Gender	54
Table 14. Items with Corrected Item-Total Correlation Less Than 0.35	56
Table 15. Descriptive Statistics of Four Semantic Modes by Gender.....	58

Table 16. MANOVA Tables by Gender for the Four Semantic

Mode Subtotal Scores.....	58-59
Table 17. Factor Structure of the Questionnaire (19 factors)	62
Table 18. Factor Structure of the Questionnaire (6 factors)	63
Table 19. Goodness-of-Fit Indices of Five Models	66
Table 20. Four Modes of Item Wording.....	84-85
Table 21. Descriptive Statistics of the Items	103-105
Table 22. Descriptive Statistics of Four modes of Item Wording	107-108
Table 23. Item-Total Statistics.....	110-111

LIST OF FIGURES

	Page
Figure 1. Histogram of Totals of 107 Items on the Questionnaire	49
Figure 2. Histogram of Mode 1 Scores on the Questionnaire.....	50
Figure 3. Histogram of Mode 2 Scores on the Questionnaire.....	50
Figure 4. Histogram of Mode 3 Scores on the Questionnaire.....	51
Figure 5. Histogram of Mode 4 Scores on the Questionnaire.....	51
Figure 6. Scree Plot from the Exploratory Factor Analysis (19 factors).....	61

CHAPTER I

STATEMENT OF THE PROBLEM

Introduction

Conventional wisdom has suggested that psychological measures should be constructed to contain an even balance of positively and negatively-worded items, so as to counteract response biases such as agreement response tendencies. The practice of using a balance of positively and negatively-phrased items in an affective instrument stems from recommendations found in the literature. Most textbooks or publications listing recommendations concerning attitude scale construction also suggest that questionnaire items should include both positively and negatively-worded item stems (Anastasi, 1982; Wiggins, 1973). Nunnally (1978) specifically advocated the reduction of response styles by having an item pool "...divided evenly between positive and negative statements" (p.605), and asserted that "stylistic variance ...can be mostly eliminated by ensuring that an instrument is constructed so that there is a balance of items keyed 'agree' and 'disagree' " (p.669). Other psychometricians have made similar statements. The general consensus in the literature has been that measures should have both positive and negative items (Scott, 1968; Anastasi, 1982).

The early data reported in support of response styles is not without challenges. Samelson (1972) stated that the researchers (Bentler, Jackson, & Messick, 1971, 1972; Couch & Keniston, 1960; Jackson, 1967a, 1967b; Jackson & Messick, 1958, 1965) failed to clarify the conceptual meaning of response styles, and used an incorrect logical model

in interpretation. The root of the problem seemed to be that all discrepant responses were defined as acquiescence – a mistake of the sort called the “the psychologist’s fallacy” (Samelson, 1972), which refers to the confusion of the researcher’s own standpoint with that of the mental fact about which he/she is making his/her report. Block (1967, 1971, & 1972) argued that acceptance (the tendency to ascribe characteristics to oneself, regardless of the direction of item keying), was not likely to be of appreciable import for understanding the nature of responses to structured personality inventories.

Rorer (1965) stated that response styles must be distinguished from response sets. He defined a response set as “the criteria according to which a respondent evaluates item content when selecting his answer,” whereas a response style was defined as “a way or a manner of responding, such as the tendency to select some particular response option independently of the item content” (Rorer, 1965, p.151). Rorer also felt it was also necessary to distinguish between achievement examinations and personality, attitude, and interest inventories, when assessing the extent to which styles affect answers to items. On examinations, but not on inventories, there were right answers and there were items whose answers the respondent must guess. Inferences concerning an individual’s response style might be made on the basis of his/her response to a number of examinations by comparing the proportion of his/her answers to any given category with the proportion keyed for that category, and by considering the proportion of wrong answers in each of the response categories. Rorer concluded that response styles were of no more than trivial importance in determining rating responses.

Despite the challenges, the practice of using positive and negative item phrasing continues to receive widespread endorsement. Psychometricians have suggested the use

of an equal number of positively and negatively-worded items as a way of reducing the possibility of a response style influencing the responses to affective instruments (Anderson, 1981; Mehrens & Lehmann, 1984). The commonly referenced, and often followed, recommendation to use an equal number of positive and negative items is based upon two assumptions. First is the assumption that the items (whether they are positively or negatively phrased) are measuring the same construct. Second is the assumption that by balancing positive and negative item phrasing, a more valid index is obtained. Although there is a wide acceptance of these assumptions, there is little research on their tenability.

In fact, the assumption that negative and positive items measure the same construct is so widely prevalent among test developers, it seems to be unquestionably accepted. The problem arises from test constructors' common practice of using negative items based on unverified assumptions. Generally, in verbal self-report measures of latent traits, it is assumed that, given standard testing conditions, an examinee's responses are determined by item content, examinee's characteristics, and, to some extent, instrument artifacts. Most Likert-type scales include a balance of semantically negative and positive-valence items, with the intent of ridding the instrument of the effects of certain response styles. It is further assumed that both negative and positive items measure the same trait. However, the empirical verification of this assumption seems to have received little attention.

Over the past several years, numerous questions have arisen pertaining to the impact of item wording on rating responses. Specifically, which item wording format is to be preferred? Is one format superior to the other, and under what constraints? How do

different modes of item wording affect rating responses? Are we measuring the same construct if we use positive and negative items? If subjects respond differently to the same item stem when the item wording format varies, could the items be regarded as nonequivalent in the same sense as content-parallel achievement items, which vary in difficulty level? More research on the effects of item wording on rating responses is needed.

Purpose of the Study

The purpose of the study was to investigate how item wording affected rating responses. The four modes of item wording: Mode 1 (regular) “I like myself,” Mode 2 (negated) “I don’t like myself,” Mode 3 (polar opposite) “I dislike myself,” and Mode 4 (negated polar opposite) “I don’t dislike myself,” were considered. Specifically, the effects of modes of item wording of rating scale items on scale and item score mean, distribution, and reliabilities were examined. The correlations between the modes of item wording and student responses were also studied.

Significance of the Study

Anderson (1981) has argued that affective characteristics facilitate desired cognitive goals of the schooling process, and are, in themselves, desired goals of the schooling process. Similarly, Bloom (1978) stated that schools should produce “independent learners” who are able to engage in higher-level thinking, develop confidence in their abilities, and possess a degree of social responsibility. In the arena of

education, increasing emphasis is being placed upon the need for valid and reliable means of assessing affective outcomes.

Most affective outcomes are currently assessed through attitude surveys. Surveys have been used widely in the measurement of attitudes and opinions. They are also a popular method for evaluating student achievement in performance-based or constructed response assessment. Furthermore, surveys are the predominant method for eliciting judgments from students on course and instructor effectiveness.

Given that surveys are so widely used in the social sciences, both as research tools and in practical applications, item development becomes an important consideration in their construction. The literature dealing with item construction is voluminous. Hundreds of articles have been published concerning issues such as the use of ratings, their reliability and validity, and potential biasing factors. Because of conflicting findings in this literature, however, it is difficult for reviewers to identify general trends.

The possible effect of item wording on overall ratings is particularly relevant to many of today's available rating scales. Yet, the current body of literature leaves numerous questions unanswered. What has yet to be determined is the possible effect of positively and negatively worded items on raters' evaluations. Do negatively worded items "encourage" a more critical evaluation than do positively worded items? Existing studies on response schemes have not directly addressed issues of general validity and the problems evolving from the specific questions stated above have not been addressed directly in the studies on response schemes which have appeared in the literature. Further research needs to be conducted to determine whether one format over another is more susceptible to rating errors of leniency or other sources of invalidity.

In survey and evaluation research, much emphasis has been placed on the development of the item stems of the questionnaire being used. Negatively worded items may highlight the negative aspects or faults of the object or person, or may serve to suggest unconsciously to the respondent particular problem areas anticipated by the researcher. If so, rating scale evaluations may be affected as much by the wording of the items as by the quality of the object or person being evaluated. The possible effect of different modes of item wording on overall ratings is particularly relevant to many of today's available student rating instruments. It has yet to be determined if the different modes of item wording influence rating responses.

Both positive and negative items are commonly used in educational and psychological measurement. Over the past several years, numerous questions have arisen pertaining to item wording. But despite the large amount of research on rating methodology, there have been relatively few conclusions concerning this measurement issue. In addition, the factors that bring some degree of control over the distributional parameters of ratings scales have received relatively little attention. There is little research on the factors that influence the meaning that subjects apply to response options when responding to rating prompts. Investigating this problem will bring some understanding of how the different modes of item wording influence the rating responses.

The underlying premise for this research is that the item wording has an influence on scale and item score mean, distribution, and reliabilities. This study makes a start toward illustrating this premise by analyzing the results derived from applying four modes of item wording to a survey. The study will also provide valuable information on an essential and important dimension of instrument development, namely, knowledge and

understanding of the effects of item wording on rating responses. The resulting information should be valuable for educators and researchers whose focus is developing effective rating scales.

This research study intends to enhance our understanding of measurement issues in item development in two ways: First, the study will lead to general conclusions on the overall relationship between item wording and rating responses. Second, the study will provide some insight about the equivalence of items between different modes of item wording. The results of the study should prove useful to administrators and faculty members who use surveys to assess affective outcomes, and to educational researchers who are looking for state-of-the art research on survey item construction.

Research Questions

The present study was designed to provide answers to the following questions:

1. What is the influence of item wording on scale and item score mean, distribution, reliabilities, and correlations between scales?
2. Are items equivalent among four different modes of item wording? The four modes are: Mode 1 (regular), "I like myself"; Mode 2 (negated), "I do not like myself"; Mode 3 (polar opposite), "I dislike myself"; and Mode 4 (negated polar opposite), "I do not dislike myself."

Overview

This chapter has presented the problem, purpose, significance, and research questions of the study. In Chapter II, a review of the literature related to the study will be presented. Chapter III describes the procedures and design of the study. The analysis and interpretation of the data is presented in Chapter IV. In Chapter V, the conclusions and implications of the study will be presented.

CHAPTER II

REVIEW OF LITERATURE

Introduction

The purpose of this study was to investigate the effects of item wording on responses on a rating scale. In developing a new survey or utilizing existing surveys, the researcher needs to examine how the wording affects subjects' responses. A number of studies have focused on various aspects of item characteristics and their possible influence on reliability and variance. Numerous researchers have also investigated the impact of item wording on rating response. Research studies were reviewed to determine the current professional opinions regarding impact of item wording on the responses of rating scale.

This chapter is divided into two major sections. The first section deals with response style. The second section focuses on item wording. Some overall conclusions follow.

Response Style

The development of scales to assess attitude poses complex methodological problems for the test constructor. Self-report, paper-and-pencil-type verbal measures are most commonly used in behavioral research and assessments. Among the various types

of verbal self-report measures, Likert-type scales are the most popular, mainly because the Likert method is conceptually simple and practically straightforward.

However, one of the major sources of criticism of self-report data centers on the susceptibility of self-report measures to various response sets that pose a continuing threat to construct validity. A good deal of research in this area was conducted in the 1950s and 1960s. Cronbach (1950) examined the effects of selected response sets on the validity of cognitive instruments, and some corrective procedures were suggested. He also identified acquiescence as a response tendency that favors affirmative responses over negative responses. Couch and Keniston (1960) called this tendency “yea-or-nay-saying,” wherein respondents consistently select in one direction, either positive or negative. The hypothesis was that some individuals have a general disposition on the positive/negative continuum regardless of the content of the items.

Various types of response sets were identified and their effects were investigated. Jacobs and Barron (1968), Green (1951), Radcliffe (1966), Stricker (1969), Wesman (1952), and Wiggins (1973) investigated the influence of social desirability in personality measurement. Couch and Keniston (1960) examined the impact of an acquiescence response set. Berg (1961) identified the deviant response set and hypothesized that it was an important dimension of personality. In fact, the literature on response styles accumulated to the point that by 1970 there had been several major reviews of literature and even reviews of the reviews (Nunnally, 1978). Bentler, Jackson, and Messick (1972), Jackson (1967a, 1967b), Jackson and Paunonen (1980), Rorer (1965), and Samelson (1972) are some of the researchers expressing their views regarding response set.

As a consequence of such an upsurge of research, controversial though it had been, it was argued that response styles do account for a certain portion of test score variance, and that if one is interested in the construct validity of the instrument, then measures should be taken to free the instrument from this stylistic variance (Nunnally, 1978, p.660). There is a relatively large body of literature on response styles which indicates that these wording changes may make a significant difference in the factor structure of scales and the item validity (Bentler, Jackson, & Messick, 1972). Bentler et al., argued this point convincingly for two different types of acquiescence response styles.

There has been considerable psychological research over the last thirty years dealing with concepts broadly classified under the general heading of “response styles.” Specifically, acquiescence response sets and various other response biases have been examined as they relate to different types of item wordings. Most of this research has utilized measures of the California F-scale (Adorno, Frenkel-Brunswik, Levinson, and Sanford, 1950), the Minnesota Multiphasic Personality Inventory (MMPI) (Hathaway and McKinley, 1967), and the Personality Research Form (PRF) (Jackson, 1967a).

Although this research has been valuable in questioning the interpretation of subject responses to these measures, it has not resolved the issue of response style relevance. The argument has been heated over whether or not response styles exist, and, if so, whether they impact upon research results in a meaningful way. Rorer (1965), for example, concluded that “the inference that response styles are an important variable in personality inventories is not warranted on the basis of the evidence now available” (p.150). Jackson and Messick (1965) responded to Rorer with extensive criticism of both his data and his conclusions. The results of their study suggested that the inclusion of

negatively worded items can result in less accurate responses and therefore impair the validity of obtained results. Thus, although the inclusion of negatively stated items may theoretically control or offset agreement response tendencies, their actual effect is to reduce response validity. This situation suggests that the current recommendation concerning the desirability of including both positive and negative items on a questionnaire may be premature and apparently warrants much further investigation.

Item Wording

Surveys are widely used in education and psychology. Because of both their widespread use and their importance, the construction of survey items has been heavily researched; yet, concerns remain about what factors influence rating responses. One of the often expressed concerns is how the item wording affects students' responses, and numerous research studies have investigated the extent to which item wording affects the rating response. By no means, though, is there total agreement on the extent of the relationship. In fact, while some investigators have found a strong relationship between item wording and rating responses, others have found no relationship at all.

Counterbalance of positive and negative items

Psychometricians recommended counterbalancing the questions which were asked, so that a positive response to one question and a negative response to another both contributed towards increasing the score on the measure as a whole (Lemon, 1973; Likert, 1932; Edwards, 1957a). This consensus has found its way to many specialty areas in educational and psychological testing. Likert (1932) suggested that those "two kinds of

statements ought to be distributed throughout the attitude test in a chance or haphazard manner.” (p. 91) Most Likert-type scales include a balance of semantically negative and positive valence items with the intent to rid the instrument of the effects of certain response styles. It is further assumed that both negative and positive items measure the same trait.

For example, Schriesheim and Kerr (1974) indicated that subject agreement response tendencies can usually be controlled by having an adequate number of negatively-worded items, and, based upon their investigation, the major existing instruments measuring perceived leader behavior were inadequate in terms of not having sufficient negative items. The authors concluded in their review that revised scales were needed, and that these measures should have a larger number of negatively worded items to offset acquiescence response biases.

A statement of caution concerning the use of negatively worded items is appropriate. Because their use may introduce covariance, some researchers have begun to question the utility of negatively worded items (e.g., Thacker, Fields, & Tetrick, 1989). Nonetheless, Schmitt and Stults (1985) suggested that covariance introduced by the direction of item wording does not necessarily result in a methodological confound that distorts conceptual interpretation. If negatively worded items are to be used, however, it would be wise to ensure that their inclusion does not present a methodological confound. It is preferable that constructs are not exclusively defined by negatively worded items during scale development. Instead, scales should have equal numbers of positively and negatively worded items.

A number of studies have focused on various aspects of scale characteristics and their possible influence on reliability, variance, and correlations. In Simpson, Rentz, and Shrum's (1976) study, questions concerning six socially-significant topics were designed. Items representing each concept were written with two criteria: strong wording versus mild wording, and positive stance versus negative stance. Each item was consequently written in four forms so as to fit in one of each possible category: "mild-positive," "mild-negative," "strong-positive," and "strong-negative."

The authors found that wording influenced responses more than the content of the items - an outcome suggesting the influence of an "agree-disagree" response set. They also found that the positive versus negative item structure influenced students' responses more than mild versus strong wording. Moreover, the authors found that the students tended to agree more with mildly worded positive items than strongly stated positive items. When reacting to the same concepts worded in the negative, students disagreed more with the strongly worded counterparts of each pair. Within some of the topical areas, students' responses were influenced more by an "agree-disagree" set than by the intended meaning of the items. Items written at a higher "emotional level" tended to elicit stronger responses when they were stated in a disagree format than when written in an agree format.

The extremity of attitude conveyed by the wording of the item also affects the mean response and variability of response. Jaroslovsky (1988) examined the effects of wording on poll results. In his study, the respondents provided different answers when the same question was asked in two different ways: "Do you think the United States should allow public speeches against democracy?" and "Do you think the United States

should forbid public speech against democracy?” The author noted that wording and context are two of the principal and closely related sources that public opinion experts refer to as “response biases” in polls. He found that even a small change in how a question was asked could trigger connotations or interpretations in the respondents’ minds that could have a major effect on how the question was answered. Moreover, he found that answers to even identical questions could vary from poll to poll, depending upon how the question is juxtaposed with others in the same survey.

The effect of wording on polls may not be obvious to many people. However, words convey tones that can have a substantial effect on the answers. In 1940, researcher Donald Rugg had pollster Elmo Roper ask similar questions of two separate national samples. The researchers found that support for free speech was 21 percentage points larger among those asked whether speeches “should be forbidden” than those who were asked whether speeches “should not be allowed.” The experiment was replicated 35 years later; researchers asked the same set of questions. The results showed a substantial increase in respondents’ willingness to tolerate free speech. People remained more willing to “not allow” speeches against democracy than to “forbid” them.

Winkler, Kanouse, and Ware (1981) examined a technique of control for acquiescence-response sets. Logical or polar opposites were used to reverse regular items. Each concept was defined by a set of matched, contradictory statements. No adverse effects of using contradictory statements were found. The authors recommended using a balance of regular and polar items for controlling acquiescence effects.

Item reversal

Researchers (Anastasi, 1982; Nunnally, 1978) suggested using an equal number of positive and negative valence statements in scales to minimize the influence of the response sets triggered by item content, the tendency to agree, and the tendency to mark in the left or right columns. However, it was difficult to determine whether the meaning of an item had been virtually reversed. Rorer (1965) investigated this aspect of item wording and concluded that many reversal pairs of items did not turn out to be reversals. Inspection of the items indicated that in many instances there was nothing at all inconsistent or contradictory about rejecting both the original and the reversed item. He indicated that while more extreme reversals result in lower correlations than less extreme reversals, the more extreme reversals are simply permitting a greater number of consistent rejections of both forms of the item.

Many studies have been conducted on the impact of reverse-scored items on survey results. In general, measurement specialists recommend a mixture of both regular and reverse-scored items in order to guard against various response biases such as acquiescence and agreement response tendency (Anastasi, 1982; Nunnally, 1978). However, there is agreement, based on intuition, that negative statements are more difficult to understand, and there is a study that suggests that negatively phrased items are less valid. Schriesheim and Hill (1981) used undergraduates to investigate the effects of positive and negative item phrasing on the validity of the responses to three forms of the same questionnaire. The authors studied the effect of item wording on questionnaire reliability and validity with data from 280 undergraduates who read a scenario describing

a hypothetical leader's behavior, and then completed one of four different questionnaires to describe that leader. The authors examined the effects of item wording on the accuracy of responses to Form XII of the Leadership Behavior Description Questionnaire (LBDQ-XII) (Stogdill, 1963). Three forms of the questionnaires (all regular items, all negated items, and a mix of the two forms of items) were randomly distributed. The responses were compared to the LBDQ-XII descriptions of a fictitious supervisor to the known levels of behavior actually shown to each subject. The fictitious supervisor was portrayed on a one-page script given to each subject. The authors found that polar opposite and negated polar opposite items had significantly lower coefficient alpha internal-consistency reliabilities as compared with those for the regular and negated regular items. Accuracy scores yielded the same results for validity of measurement. The authors also found that regular wording format yielded more accurate responses than the negated or mixed formats. The lowest level of respondent accuracy was found in the mixed format.

Chang (1995) examined the psychometric equivalence of negative and positive items. Some researchers call negative items "semantically negative," meaning they have a negative meaning. Such a definition lacks accuracy. "Semantic" refers to the formal meaning, or the nature of a statement free from value judgment or sentiment. The value judgment or feeling of a word or statement is represented by its connotation. For example, the words frugal and miserly have the same formal meaning, whereas one has a positive or neutral connotation and the other has a negative connotation. Similarly, a test item can be connoted as positive or negative, whereas its semantics are free from a positive or negative sentiment. Chang suggested that defining items as semantically negative is incorrect. Items can be defined not in terms of the manifest syntax or

semantics but in terms of the underlying connotation. The opposite connotations of items represent two directions of a latent construct continuum of which items or their semantics are indicators.

Chang defined a test item as connotatively consistent or connotatively inconsistent when the connotation of the item either agrees (conforms with) or disagree (contradicts) that shared by the majority of the items making up a test or a subscale of a test. He examined the equivalence between connotatively consistent and connotatively inconsistent items of a 4-point Likert type scale using confirmatory factor analysis. His study concluded that these items measured correlated but distinct traits. He also suggested that connotatively inconsistent items should not be used. Items on a test or questionnaire should possess items connotatively consistent with the construct being measured.

Ahlawat (1985) concluded that semantically negative and positive item contents do not measure essentially the same construct. His study was based on a sample of Jordanian middle school students using an Arabic translation of State-Trait Anxiety Inventory (Spielberger et al., 1970). Four sets of items were constructed with distinct modes of semantic presentation. He suggested that double negative items create cognitive complexity for the students, which may end up in confusion. On the basis of the correlational and variance-related analyses, his findings concluded that semantically negative and positive item contents do not measure essentially the same construct. Furthermore, the author suggested that more cognitive steps are required to decode or unravel a negatively worded statement, thus making the task more difficult than responding to a positively worded statement. The findings of this study question the

common practice of using both negative and positive valence items in a scale and making decisions on the basis of the students' total score on items. This problem deserves closer inspection through more specifically designed studies.

Benson and Hocevar (1985) examined the effect of item phrasing on the validity of Likert-type attitude scales. Three content-parallel forms of the same questionnaire were developed to assess student attitudes toward integration. The forms differed only in terms of item phrasing. The first contained 15 positively phrased items. The second form contained the same 15 items phrased negatively. The third form contained the same 15 items, eight with positive phrasing, and seven with negative phrasing. The words "was not" were used to create a negative statement from the positive statement.

The study reported strong evidence that the insertion of the word "not" has a profound influence on student responses. The items that induced a favorable response on the positive form induced a less favorable response on the negative form. Respondents were less likely to indicate agreement by disagreeing with a negatively phrased item than to indicate agreement by agreeing with a positively phrased item. Moreover, items that induced an unfavorable response on the positive form were less likely to induce an unfavorable response on the negative form. The analyses provided evidence that changing positive statements into negative statements may have an effect on the psychometric characteristics of an item. These analyses did not provide conclusive evidence that positive and negative items measure the same construct in different ways. The results indicated that the subjects had difficulty expressing agreement by disagreeing with the negated items.

In order to examine if a different construct was being measured, two models were contrasted. One model composed of two factors where the factors were set to be correlated (undifferentiated model), and a second model where the correlation between the two factors was estimated (differentiated model). The authors found that the differentiated, two-factor model provided a better fit to the data than the undifferentiated two-factor model. It was suggested that the positive to negative transformations change not only an item's psychometric characteristics, but also change the construct that an item is intended to measure.

Campbell and Grissom (1979) examined the effect of item phrasing on attitude scale items. Two forms were developed – the first containing all regularly-scored items and the second consisting of items that were designed to be their logical opposites (polar). The results of factor analyses suggested that the two different formats measured different constructs. The authors also indicated that scoring-negated or polar-attitude scale items were not equivalent to the reversal of regular items.

Schmitt and Stult (1985) suspected that a small number of respondents who were careless in their responses might be responsible for the appearance of negative factors composed only of reverse-scored items. The objective of their study was to show how a “negative factor” can be produced by a relatively small number of careless respondents. A frequently occurring phenomenon in the analysis of personality or attitude scale items is that all or nearly all questionnaire items that are negatively keyed define a single factor. Although substantive interpretations of these negative factors are usually attempted, this study demonstrated that the negative factor could be produced by a relatively small portion of the respondents who failed to attend to the negative-positive wording of the

items, and who did not notice that some items are the opposite in meaning to the majority of the items. In a series of simulations, the proportion of “careless” respondents and the proportion of negatively keyed items were varied for data generated from three different correlation matrices reflecting different levels of item intercorrelation. The results indicated that when only a small portion of the respondents were careless (ten percent), a clearly definable negative factor was generated. The authors cautioned about the use of item reversal and the interpretation of the reversed-score items.

In summary, it was concluded that scales with negatively keyed items frequently led to the identification of a factor defined wholly or mostly by those negatively keyed items. Other literature cited above indicates that this finding is relatively widespread in the sense that it occurs in a variety of research areas.

Marsh (1984), who found a negative item factor in a self-concept measure for elementary school children reported similar results. Other researchers have also reported negative factors (Campbell & Grissom, 1979; Simpson et al., 1976), thus adding support to the notion that negative phrasing may actually change the construct that the item is intended to measure.

Harasym (1992) reported, from a study with approximately 200 first-year nursing students, evidence that the use of negation (e.g., not, except) should be limited in stems of multiple-choice test items, and that a single-response, negatively worded item should be converted to a multiple-response, positively worded item.

Several other researchers (Andrich, 1983; Campbell & Grissom, 1979; Simpson, Rentz, & Shrum, 1976) have investigated whether phrasing can influence overall attitude levels on different attitudinal questionnaires. These researchers all concluded that item

phrasing makes a difference. However, the results these researchers report cannot be easily corroborated with each other. At issue is that the word “not” was used in Andrich’s study (1983) to create parallel negative statements, whereas the other researchers cited created negative statements by item reversal (Campbell & Grissom, 1979; Simpson, Rentz, & Shrum, 1976). Rorer (1965) suggested that this latter procedure often leads to negative statements that reflect different content or ideas; consequently, such statements are not direct opposites of the original positive statements. It is because of this problem that many affective scales contain the word “not” to create a negative statement (Coopersmith, 1967; Marsh, Smith, Barnes, & Butler, 1983; Piers, 1969).

The findings that relate to the present study are summarized in Table 1. For each review, the table shows the name of investigator, year of publication, and a summary of findings.

Table 1. Summary of Findings Related to the Study.

Investigator	Year	Finding
1. Likert Edwards Lemon	1932 & 1957 & 1973	The authors recommended counterbalancing the survey items so that a positive response to one question and a negative response to another both contributed towards the score on the measure.
2. Cronbach	1946 & 1950	There is a tendency for people to favor affirmative responses over negative responses.
3. Couch & Keniston	1960	The authors labeled the tendency wherein respondents consistently select in one direction, either positive or negative, as "yea-or-nay saying."
4. Rorer	1965	Many reversal pairs of items do not turn out to be reversals.
5. Simpson, Renz, & Shrum	1976	Wording influenced responses more than the content of the items.
6. Campbell & Grissom	1979	Scoring-negated or polar-attitude scale items are not equivalent to the reversal of regular items.
7. Schriesheim & Hill	1981	Regular wording format yielded more accurate responses than the negated or mixed formats. The lowest level of respondent accuracy can be found in the mixed format.
8. Winkler, Kanouse & Ware	1981	No adverse effects of using contradictory statement was found. The authors recommended using a balance of regular and polar items for controlling acquiescence effects.
9. Ahlawat	1985	Semantically negative and positive item contents do not measure essentially the same construct.

Table 1. (cont'd)

Investigator	Year	Finding
10. Benson & Hocevar	1985	A detrimental effect would occur if the number of regular and reverse-score statements were balanced.
11. Schmitt & Stult	1985	A frequently occurring phenomenon in factor and cluster analysis of personality or attitude scale items is that all or nearly all questionnaire items that are negatively keyed will define a single factor. The study demonstrates that the negative factor could be produced by a portion of respondents who fail to attend to the negative-positive wording of the items.
12. Jaroslovsky	1988	The wording of the item affects the mean response and variability of response.
13. Harasym	1992	A single-response, negatively worded item should be converted to a multiple-response, positively worded item.
14. Chang	1995	Connotatively consistent and connotatively inconsistent items measured correlated but distinct traits. Items on a test or questionnaire should be connotatively consistent with the construct being measured.

Summary

Surveys are used extensively in a wide range of assessment. The increase in popularity of surveys as measures of affective outcomes consequently has focused a great deal of attention on their validity. It is important to understand the ways in which people use survey items and, especially, to understand the factors that can influence the responses given. Many researchers have provided an extensive review of research in this area. Previous research has not addressed the perceived defensibility or accuracy of the

assumption of construct equivalence with regard to using both positive and negative items in a survey.

The literature review has provided insights into the effects of response format on rating scales. The various research studies suggest that item wording does account for a certain portion of test score variance. If we are interested in the construct validity of the instrument, then measures should be taken to account for this stylistic variance.

(Nunnally, 1978, p.660). However, the cited studies have shown inconsistent findings with respect to the effects of wording of items on the rating scale. The desirability of including a mixture of regular and reversed-scored items on attitude and questionnaire measuring instruments is yet to be determined conclusively. Research studies yielded inconsistent and ambiguous support for balancing regular and reversed-scored items (Bentler, Jackson, and Messick, 1972; Jackson and Messick, 1965; Rorer, 1965). This circumstance raises serious concern about whether both regular and reversed-scored items should be included on a measuring instrument. The research results of these studies indicate that further investigation is needed.

Although the recommendations of some authors are in conformity with conventional psychometric advice, the experience of the authors suggests that the use of negative items may have at least some dysfunctional consequences. In their experience, negatively worded items often reduce scale reliability, and they may be eliciting response biases or measuring unintended aspects of constructs under investigation. In any event, as measurement validity requires instrument reliability, these impressions suggest that the use of negative items may not be cost-free.

The investigators have raised serious questions about the influence of item wording on student responses. Because of conflicting findings, however, it is difficult to draw firm conclusions. Furthermore, several complexities in this research literature add to the difficulty of drawing overall generalizations concerning the impact of item wording. The research to date suggests that positive to negative transformations change an item's psychometric characteristics and, more importantly, change the construct that an item is intended to measure. However, the studies that have been reviewed do not show that positively phrased items are necessarily better indicators of attitude. Nevertheless, there is some indication that negatively phrased items are less valid. There is the plausible argument that respondents may not understand that they can indicate agreement by agreeing with a negative statement. Marsh (1984) provided support for the above contention. Despite the conventional wisdom so often found in measurement textbooks, pronouncements by researchers in the study of item phrasing have been unanimous that negatively phrased items reduce the validity of a questionnaire (Andrich, 1983; Campbell & Grissom, 1979; Marsh, 1984; Schriesheim & Hill, 1981; Simpson et al., 1976).

One measure almost universally adopted in Likert-type scales to minimize the influence of the response set triggered by item content, the tendency to agree, and the tendency to mark in the left or right columns is to include an equal number of positively and negatively valenced statements in the scales. An actual reversal of meanings of an item, however, may be hard to achieve. To complicate matters still further, it is not easy to determine that the meaning of an item has been virtually reversed. Rorer (1965) has provided many examples of reversed pairs of items that on close scrutiny turn out not to be reversals.

The assumption that negative and positive items measure the same construct is so widely prevalent among test developers that it seems to be unquestionably accepted by almost everybody. Test constructors' conviction in this assumption is further fortified by empirically obtained high indices of homogeneity contrary to the psychometricians' warning that homogeneity neither implies nor guarantees the unidimensionality of the trait being measured by the test.

The problem arises out of the test constructors' common practice of using negative items based on unverified assumptions. Generally, in verbal self-report measures of latent traits, it is assumed that, given standard testing conditions, an examinee's response is determined by item contents, examinee characteristics, and, to some extent, instrument artifacts. Most Likert-type scales include a balance of semantically negative- and positive-valence items with the intent of ridding the instrument of the effects of certain response styles. It is further assumed that both negative and positive items measure the same trait. The empirical verification of this assumption seems to have received little attention from researchers.

The purpose of this study was to investigate the effects of different modes of item presentation on the responses of various groups of high school students. Taking a simple case of semantic change, a sentence with a positive import, for example, "I like myself," can be reversed in meaning either by replacing the word "like" with one of its antonyms (e.g., dislike) or by using a grammatically negative mode (e.g., "I do not like myself"). Similarly, a negatively valenced sentence can be reversed in meaning by using the same transformations. Following this scheme, four distinct modes of semantic presentation can easily be defined: Mode 1 (regular), "I like myself;" Mode 2 (negated), "I do not like

myself,” Mode 3 (polar opposite), “I dislike myself”; and Mode 4 (negated polar opposite), “I do not dislike myself.”

More specifically, the study looks into the following questions: What is the influence of wording of rating scale items on rating scales? What is the influence of item wording on scale and item score mean, distribution, reliabilities, and correlations between scales? What is the item equivalence between different modes of item wording? Do different modes of presenting the seemingly same content measure the same trait? Does the grammatically negative mode incorporate ambiguity into the items? How much variance is due to format factors? Does stating a concept in a positive manner affect reactions differently than stating the same concept in a negative manner? Do students agree with positively stated items to the same extent they disagree with the concept when stated negatively? These questions warrant further investigation of the effects of item wording on rating responses.

CHAPTER III

RESEARCH DESIGN AND PROCEDURE

Introduction

The purpose of this study was to investigate the effects of item wording on the rating responses. This chapter includes a description of the development of the test instrument, selection of the sample, procedures of test administration, quality control screening, and the statistical procedures for examining the research questions of this study.

Development of the Test Instrument

General Shame Scale

The General Shame Scale was developed by Chang & Hunter (1988). On the basis of an examination of major writings in shame literature, everyday language, and clients' verbal reports, the authors developed and tested scales measuring each of six shame themes. They are disappointment with oneself, feelings of inferiority, feelings of defectiveness, feelings of worthlessness, feelings of unimportance, and feelings of falling short of one's own standards and ideals. Items were written to avoid reference to other affects such as embarrassment or self-consciousness. These shame themes were highly correlated and were shown by confirmatory factor analysis to measure one underlying factor. The specific factors were shown to be uncorrelated with measures of emotional and social function once the general shame was partialled out. They then combined the

shame theme scales to form a general shame scale. These shame themes were developed into six shame theme scales. Initially, 65 items were written altogether for six shame theme scales. Mainly on the basis of content meanings, 22 best items were retained. The reliability of the 22 item general shame scale was .95. The shame theme scales are presented in Table 2.

For this study, four versions of the items in four distinct modes of semantics were prepared (Appendix A). In the first version, the statements were presented by semantically positive words or phrases. For example, “ I like myself.” In the second version, the statements were presented by semantically positive words or phrases structured in grammatically negative sentences. The 22 sentences of the first version were transformed into “do not” sentences, for example, “ I do not like myself.” In the third version, “I dislike myself,” the statement was reversed in meaning by replacing the word “like” with one of its antonyms (e.g., dislike). Similarly, in the fourth version, the items in the third version were transformed into “do not” sentences, for example, “I do not dislike myself.” Since English is not the first language of the author of this study, the items were carefully reviewed by a native speaker to ensure the accuracy of the conversion of the four modes of item wording.

Chang & Hunter (1988) have shown that shame relates very substantially to emotional and social problems. In their study, people high in shame are high in anxiety ($r=0.86$) and low in life satisfaction ($r=-0.73$). Hence, for the sake of validation, two other scales, which measure anxiety and life satisfaction, were used in this study.

Dimensions of Self-Concept (DOSC)

The Dimensions of Self-Concept (DOSC) (Michael & Smith, 1976) was developed to measure non-cognitive factors associated with self-concept in a school setting. There are two main purposes for the DOSC: a) to identify those students who might experience difficulty in their schoolwork because of their perceptions of a low degree of self-esteem; and 2) to diagnose for purposes of counseling or guidance by the teacher, professional counselor, or administrator those dimensions as well as the specific activities associated with them that might contribute to low self-esteem and might impair learning capabilities relative to negative affectivity.

The DOSC is a self-report instrument that reflects the perception that students have regarding each statement of the five main dimensions: aspiration, anxiety, academic interest and satisfaction, leadership and initiative, and identification vs. alienation. The five factor dimensions measured by the DOSC scales are described as follows:

1. Level of Aspiration

This factor is a manifestation of patterns of behavior that portray the degree to which achievement levels and academic activities of students are consistent with their perceptions of their potentials in terms of scholastic aptitude or past and current attainments.

2. Anxiety

This factor reflects behavior patterns and perceptions associated with emotional instability, a lack of objectivity, and a heightened concern about tests and the preservation of self-esteem in relation to academic performance.

3. Academic Interest and Satisfaction

This factor portrays the love of learning and pleasure gained by students in doing academic work and in studying new subject matter.

4. Leadership and Initiative

This factor appears to represent those behavior patterns and perceptions that are associated with star-like qualities, in which a student has an opportunity to demonstrate his mastery of knowledge, to help others, to give direction to group activities, to become a respected expert whom others consult, and to put forth sound suggestions for classroom activities.

5. Identification vs. Alienation

This factor is intended to represent the extent to which a student feels that he has been accepted as part of the academic community, has been regarded by his teachers, and peers as a significant person.

The 14-item Anxiety Factor was used in this study (Table 3). A 0.82 coefficient alpha was reported for the 14-item scale (Michael & Smith, 1976).

Satisfaction with Life Scale

The Satisfaction with Life Scale (Diener et al., 1985) is a five-item scale, which measures global life satisfaction. The scale is designed to assess global life satisfaction and does not tap related constructs such as positive affect or loneliness. The purpose of the scale is to obtain an overall judgement of the respondent's life in order to measure the of concept of life satisfaction. In the initial phase of item construction, 48 self-report

items were generated. A five-item scale was formed after a factor analysis and an examination of the semantics. Diener et al. (1985) reported a high reliability (0.87 coefficient alpha) for the scale. It also correlates moderately to highly with other measures of subjective well-being. The scale is well suited for use with different age groups. The high correlations with personality indicators of well-being suggested that the scale might also be useful in clinical settings (Table 4).

Questionnaire Development

The 107 items were constructed and translated into equivalent Chinese structures. The preservation of the connotation of meaning and the essence of the original sentences was the top priority. Furthermore, the items (in Chinese) were edited for clarity and readability by school teachers, who were requested to edit each item with respect to its clarity and suitability for the middle and high school students. The Chinese version of the questionnaire was refined on the basis of the teachers' comments. When the English version of the questionnaire was given to the Chinese teacher who assisted with the translation, she raised a concern regarding the rating scale. In the English version, the anchors of the rating scale are "Never," "Seldom," "Sometimes," "Often," and "Almost Always." She reminded the author that the last scale anchor "Almost Always" might not be an appropriate word choice. In the Chinese version "Almost Always" was changed to "Always."

In order to ensure the appropriate wording of the survey and the students' understanding of the survey, a pilot test was conducted in two classrooms of about 80 to 90 students. The subsequent item analysis served as a basis for further refinement of the

translation and wording of the survey. Based on the recommendations of the teachers and the students, the researcher made some very slight changes to the wording of some the survey items. The items were then assembled in random order into a questionnaire with instructions. A copy of the survey can be found in Appendix B. The survey in Chinese can be found in Appendix C. Subjects also completed items on their gender and age.

Table 2. The General Shame Scale (Chang & Hunter, 1988)

(1) Disappointment with Oneself

1. I don't like myself.
2. I am pleased with myself.
3. I am disappointed in myself.
4. I feel ashamed of myself.

(2) Feelings of Inferiority

5. I feel like I am just not quite good enough.
6. I feel that I am inferior to most of my friends.
7. Compared to others, I feel like I don't measure up.
8. I am just as good as my friends.

(3) Feelings of Defectiveness

9. I feel inadequate as a person.
10. I feel there is something defective in my character.
11. I look down on myself because of my flawed character.
12. I see myself as intact and without personal defects.

(4) Feelings of Worthlessness

13. I feel worthless as a person.
14. I feel like a useless person.
15. I feel like I am good for nothing.
16. I feel I am a complete failure as a person.
17. I feel like a failure.
18. I am a worthwhile person.

(5) Feelings of Unimportance

19. I feel unimportant.
20. I feel so insignificant to others, as if I were invisible.

(6) Falling Short of Own Standards and Ideals

21. I always seem to fall short of my aspirations.
22. I find that I don't live up to my own standards and ideals.

Table 3. Dimensions of Self-Concept (DOSC) (Anxiety Factor)
(Michael & Smith, 1976)

1. Statements that some teachers make about my schoolwork hurt my feelings.
2. I feel so nervous about some of my classes that it is hard for me to attend.
3. I become tense and nervous when I am studying.
4. I am upset about so many things that I cannot concentrate on or do my schoolwork.
5. I worry about how well I am doing in my classes.
6. I am afraid to ask teachers to explain a difficult concept a second or third time.
7. I avoid talking to my classmates about schoolwork because they might make fun of me.
8. I become frightened when a teacher calls on me in class.
9. Talking in front of class makes me feel nervous.
10. I feel upset when I have to take a test.
11. I would be afraid to tell a teacher that he or she made a mistake in explaining an assignment or in working a problem.
12. I have trouble sleeping well the night before an important examination.
13. I am embarrassed to face my friends or family if I have made a low grade on a test or assignment.
14. I worry that my score on a test will not be one of the highest in class.

Table 4. Satisfaction with Life Scale (Diener et al., 1985)

1. I am satisfied with my life.
2. The conditions of my life are excellent.
3. In most ways, my life is close to my ideal.
4. So far, I have gotten the important things I want in life.
5. If I could live my life over, I would change almost nothing.

Selection of the Sample

Participation in the study was solicited from students enrolled in middle schools. The subjects in this study were the students studying in the sixth and seventh grades in two schools located in Taipei, the capital of Taiwan. Although the schools were not randomly selected, they were considered typical in terms of achievement, socioeconomic status, and ethnic background to other schools within the district. In Taiwan, there are usually 35 to 40 students in one classroom. Twenty-two classrooms were sampled.

Procedures of Test Administration

The questionnaires were administrated during class sessions. No titles were printed on the questionnaires. The questionnaire required approximately 20 minutes to complete. Before responding, students were told only that the survey dealt with opinions about themselves. Subjects were instructed to use all five points on the rating scale to provide an accurate reflection of their opinions. Subjects were encouraged to respond to all items. The teacher read the survey instructions aloud as the students followed along. Students were told that their responses would be completely anonymous. No names or other identifying information were collected. Students were told that the survey was completely voluntary, that they did not have to participate, and that they could leave unanswered any questions they thought were too personal. The teachers left their classrooms while the questionnaires were administered.

Quality Control Screening

Quality control procedures were developed with prior surveys to screen for incomplete or otherwise unusable responses. Students were instructed to respond to each item of the questionnaire. After the questionnaires were completed, the students' responses were entered into an SPSS data file. A printout of this file was obtained and each of the entries was then checked for mis-entry with each of the students' surveys. All missing responses were coded "9".

Each questionnaire was carefully inspected for any detectable abnormalities. Such aberrant cases were discarded from the sample. A screening procedure was applied to exclude students who were not taking the survey seriously. If the student marked one particular option constantly throughout the survey, that survey was discarded. A questionnaire was considered unusable if the student did not answer any items.

Data Analyses

Items representing each mode were summed to obtain a subtotal score for each mode. Total test score was obtained by adding the four subtotal scores. It was reasoned that if subjects' responses were mainly determined by item contents, as hypothesized, then all four categories of items would tap the same source trait (construct), and consequently, the correlation coefficients among the pairs of mode subtotal scores would be high; otherwise, they would be low. The Mode 1 (regular - "I like myself") and Mode 4 (negated polar opposite - "I do not dislike myself") variables were assumed to be positive, and the Mode 2 (negated - "I do not like myself") and Mode 3 (polar opposite -

“I dislike myself”) were assumed to be negative aspects of the same construct. The intermode Pearson product moment correlation coefficients were computed.

Scale responses were scored by reversing all negatively worded questionnaire statements. Items representing each version were summed to obtain a subtotal score for each mode. A higher score on any of the subscales or on the total scale indicated a more positive attitude toward a certain aspect. The means, standard deviations, and reliabilities of the six-shame theme scale and the general shame scale were computed. The intercorrelations between six shame themes were also computed.

Hunter and Gerbing (1982) noted that if the right research design is used, then confirmatory factor analysis can be used to assess items using the criteria: "internal consistency" and "parallelism" (or "external equivalence"). If item responses differ from each other only by random error of measurement, then the item errors will not correlate with each other. The correlations between items within a scale should then satisfy a mathematical product rule discovered by Spearman (1904, cited in Hunter & Gerbing, 1982) – his “one factor model.” If the correlations between items within a scale satisfy the Spearman product rule, then the scale is said to be "internally consistent." This is a weak criterion for item equivalence.

Hunter and Gerbing (1982) also noted that there is a stronger criterion for item equivalence – parallelism in the pattern of correlations between the items and important "outside" variables, such as the measures of emotional and social functioning used in the present study. If all items measure the same construct, then the item errors will not correlate with outside variables. The correlations between items in a unidimensional scale with any outside variable should satisfy a condition called "parallelism" (Tryon,

1939; Tryon & Bailey, 1970; Hunter, 1973; Hunter & Gerbing, 1982). This is a strong test for item equivalence.

Confirmatory Factor Analysis

A confirmatory factor analysis was run on the items organized into the predicted four clusters (modes of item wording). The analysis was used to examine the quality of each scale. Confirmatory factor analysis is a method for computing the correlations for constructs from the correlations for observed measures, so long as the measures obey the assumed measurement model. Each scale was checked for homogeneity of content, for internal consistency, and for external consistency or parallelism. The first analysis was done to assess each of the four scales separately. If there was no significant departure in fit, each of the four scales would prove internally consistent. The items within scales would be parallel in relationship to the other scales and to that extent would appear to be equivalent to each other.

A hierarchical measurement model was constructed with regard to the relationship between the modes of item wording and the responses. A hierarchical confirmatory analysis estimated the correlations between the item scores and the four modes of item wording. The correlations between the four scales and each of the measures of anxiety and life satisfaction were also computed. A confirmatory factor analysis of the four modes of item wording together fits the data only if each scale is parallel to each other scale in their pattern of correlations with them. A stronger test of parallelism was obtained by testing the items in each scale for parallelism in terms of how the items

related to outside variables. In order to do this, each mode of item wording was tested separately for parallelism in relationship to the measures of anxiety and life satisfaction.

If the first stage of analysis shows that each of the scales is unidimensional, then a second stage of analysis can test the hypothesis that the four modes of item wording are each measures of one underlying trait. This analysis first checks to see if the separate scales define specific factors or if they are identical to each other. If the scales are not highly correlated with each other, then there are specific factors, which differentiate the shame constructs from one another.

Exploratory Factor Analysis

To see if there might be some completely unanticipated dimension in the data, an exploratory factor analysis of the items was also run. The communalities were estimated as the largest correlation. The principal axis factors were followed by VARIMAX rotation. The eigenvalue cutoff for the number of factors was set at 1.00. An examination of the resulting factors would show if the clusters matched with the a priori clusters. In other words, when the items were blindly grouped together using the highest loading from the varimax factors, the clusters thus formed should be similar to the original clusters (modes of item wording).

Parallelism

Parallelism is the basis for the use of correction for attenuation to eliminate the bias in correlations produced by error of measurement. Thus the test for parallelism is the test which directly justifies the use of correction formulas. Since the correction formulas are implicit in confirmatory factor analysis, it is the test for parallelism that is the heart of the assumptions for confirmatory factor analysis.

Items in each of the four scales were examined in terms of how they correlated with outside variables. If the items in a cluster all measure the same affect, then the items in that cluster should correlate in a parallel way with each of the outside variables. Parallelism was tested by computing the correlations between the items and the outside variables (anxiety and life satisfaction), and by examining that correlation matrix for parallelism (as noted by Hunter, 1973, and Tryon and Bailey, 1970). The most direct way to check for equivalence is to correlate scales measuring the same thing (though it is important to correct for the attenuation produced by random error of measurement). If two measures are equivalent, then they correlate in exactly the same way with other variables. The correlations between all four scales, anxiety (DOSC Anxiety Factor) and life satisfaction (Satisfaction with Life Scale) were computed and were corrected for attenuation due to error of measurement.

Summary

The research design of this study was aimed to address two main issues concerning the effect of item word in the rating responses. The first issue was the influence of item wording on scale and item score mean, distribution, reliabilities, and

correlations between scales. The second issue was the item equivalence between different modes of item wording. Two Taiwanese schools were selected for this purpose, all from an urban setting in Taipei. A questionnaire was constructed and administered to a total of 861 sixth and seventh grade students from the four schools. Demographic information was also obtained regarding the students' grade, gender, and age. All this information was coded into a SPSS file, cleaned and subjected to a data analyses. Descriptive statistics, correlational analyses, and reliability analyses were obtained. A confirmatory factor analysis was also conducted to further analyze the data. Various other statistical tests were also performed on the results of the survey data.

In summary, the findings of several other researchers clearly suggest that item wording can make a difference. They have not shown that a different construct was being measured. This study provides a second set of analyses to test whether altering item wording had an effect on the nature of the construct being measured. The results will have important implications for measurement instrument design by addressing the question: just what are the effects of item wording on the item responses? The purpose of this current research was, therefore, to explore the item wording issue further and to clarify its implications for the validity of questionnaires.

CHAPTER IV

ANALYSES AND INTERPRETATION OF THE DATA

Introduction

This chapter presents the data analyses. First, a general description of the characteristics of the sample will first be presented. This will be followed by an account of the factor analyses on the questionnaire data. The manner in which the analyses were conducted will be explained and its implication on survey development based on the results will be discussed. Finally the results for the two research questions of the study together with their interpretations will be reported.

Characteristics of the Sample

A total of 861 surveys were returned. All the surveys were carefully inspected before data entry. Eleven surveys were discarded because participants marked the same score on each item. One student did not fill out the survey.

A total of 849 useable surveys were included in this study. They were completed by students from two middle schools in Taipei. There were 443 (52.2%) male and 406 (47.8%) female students. All students were from the urban setting. Their age ranged from 12 to 17. Eleven (1.3%) students were twelve years old; 152 (17.9%) were thirteen; 451 (53.1%) were fourteen; 228 (26.9%) were fifteen; 5 (0.6%) were sixteen; and

2 (0.2%) were seventeen. Two hundred and fifty-six students (30.3%) were in the sixth grade and 593 students (69.8%) were in the seventh grade. Table 5, 6, 7, and 8 show the distribution of students by gender, age, and grade.

Table 5. Gender of the Participants

	Frequency	Percent
Male	443	52.2%
Female	406	47.8%
Total	849	100.0%

Table 6. Age of the Participants

Age	Frequency	Percent
12	11	1.3%
13	152	17.9%
14	451	53.1%
15	228	26.9%
16	5	0.6%
17	2	0.2%
Total	849	100.0%

Table 7. Grade Level of the Participants

Grade	Frequency	Percent
6 th Grade	256	30.3%
7 th Grade	593	69.8%

Table 8. Participants' Age and Gender by Grade

	6 th Grade	7 th Grade
Gender		
Male	131	312
Female	125	281
Age		
12	11	0
13	132	20
14	104	347
15	9	219
16	0	5
17	0	2

Analyses of the Questionnaire Data

The 107 items were coded into six subscales as shown in Table 9. The four modes of the shame scale represented four different modes of semantics for a total of 88 items (as mentioned in Chapter 3.) In Mode 1, the statements were presented by semantically positive words or phrases. In Mode 2, the statements were presented by semantically positive words or phrases structured in grammatically negative sentences. The 22 sentences of Mode 1 were transformed into “do not” sentences. In Mode 3, the sentences were reversed in meaning by replacing each adjective with one of its antonyms. Similarly, in Mode 4, the items in Mode 3 were transformed into “do not” sentences. DOSC-Anxiety is a 14-item scale, which measured student academic anxiety. Satisfaction with Life scale is a 5-item scale, which measured global life satisfaction.

Table 9. Coding Format of the Subscales in the Questionnaire

Modes	Question Number
Mode 1	2, 3, 9, 13, 24, 25, 36, 41, 48, 49, 57, 61, 67, 81, 83, 85, 87, 88, 95, 100, 102, 104
Mode 2	4, 10, 20, 23, 31, 34, 43, 46, 51, 59, 60, 64, 72, 78, 82, 86, 89, 93, 96, 99, 101, 105
Mode 3	7, 8, 12, 18, 30, 32, 35, 37, 39, 50, 52, 54, 55, 66, 73, 75, 77, 80, 84, 92, 94, 97
Mode 4	1, 5, 6, 14, 16, 17, 21, 26, 27, 40, 42, 44, 47, 53, 56, 63, 68, 71, 79, 91, 98, 107
DOSC-Anxiety Factor	11, 15, 19, 29, 38, 45, 62, 65, 69, 70, 74, 90, 103, 106
Satisfaction with Life Scale	22, 28, 33, 58, 76

Answers to the Research Questions

Research Question 1: What is the influence of item wording on scale and item score mean, distribution, reliabilities, and correlations between scales?

Histogram Analyses

To determine if the sample chosen was representative of a normal population, a histogram of the student's scores on the questionnaire was plotted (see Figure 1). The mean and standard deviation of the 107-item questionnaire were 298.2 and 53.52 respectively (Appendix D). The item descriptive statistics of the four modes can be found in Appendix E.

The plot showed that the distribution of the total scores was not normally distributed. The test of normality was significant at $p < 0.0001$ level. For purposes of comparison, the points of a normal curve base on all valid values of the scores were superimposed on the histogram.

Four other histograms were plotted for the four modes of items in the questionnaire (see Figures 2, 3, 4, and 5). They were also found to be not normally distributed, except Mode 1 (Table 10). The means of the four Modes were 68.5 (Mode 1), 78.7 (Mode 2), 83.1 (Mode 3), and 67.9 (Mode 4). The means of Mode 2 and Mode 3 were higher than Mode 1 and Mode 4.

Table 10. Tests of Normality

	Kolmogorov-Smirnov	
	Statistic	Sig.
Total of 4 Modes	.047	.000
Mode 1	.028	.142
Mode 2	.053	.0001
Mode 3	.084	.0001
Mode 4	.043	.001

Figure 1. Histogram of totals of 107 items on the questionnaire.

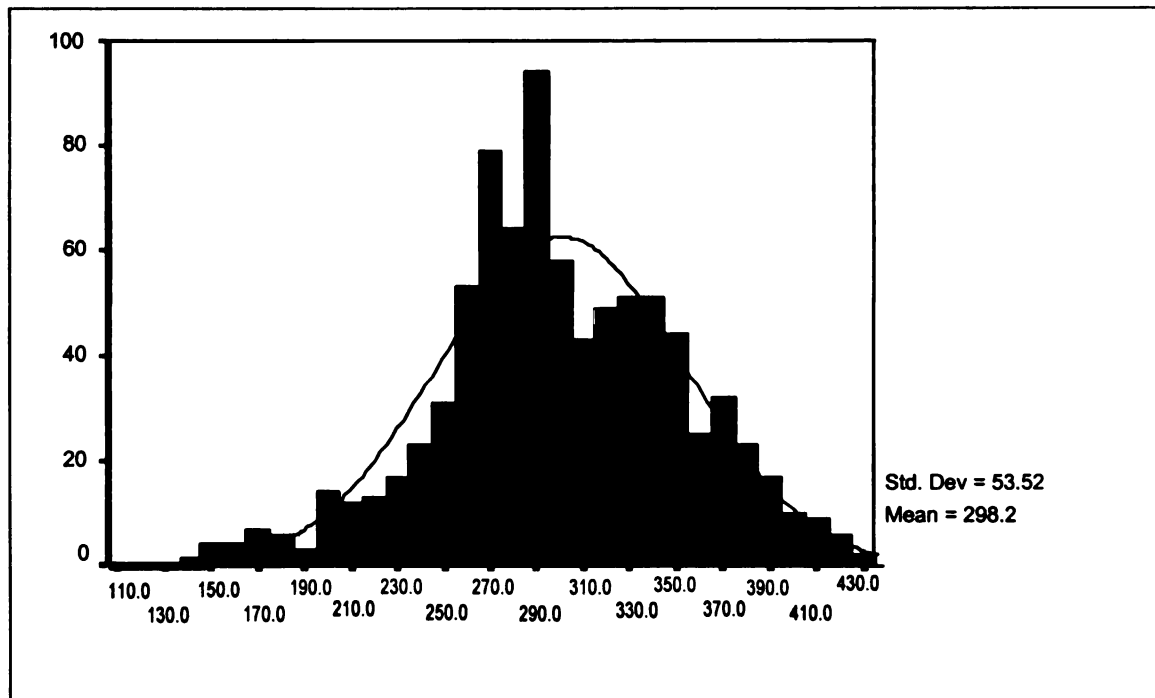


Figure 2. Histogram of Mode 1 scores on the questionnaire

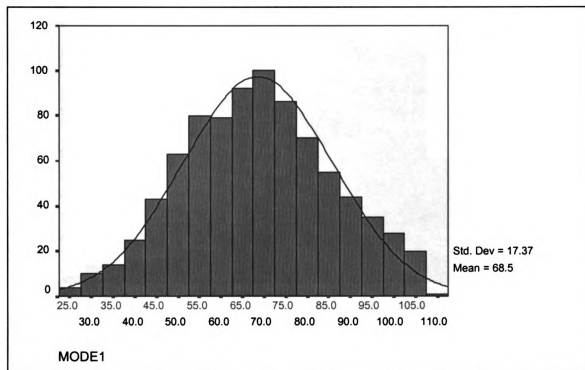


Figure 3. Histogram of Mode 2 scores on the questionnaire

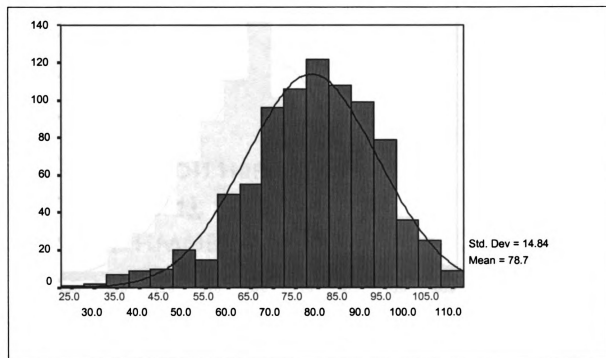


Figure 4. Histogram of Mode 3 scores on the questionnaire

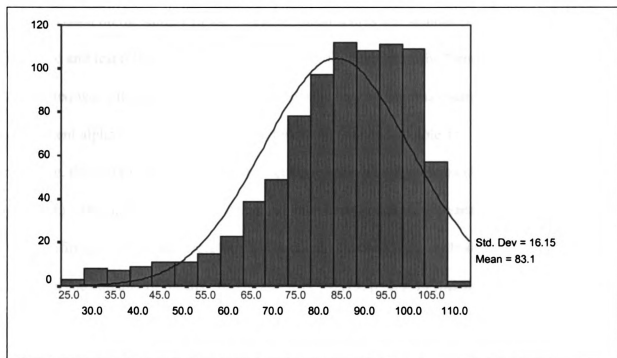
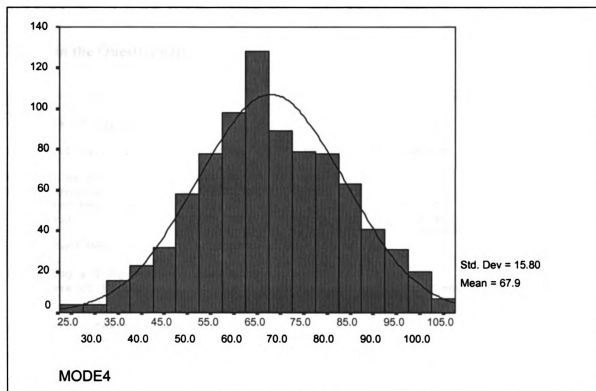


Figure 5. Histogram of Mode 4 scores on the questionnaire



Reliability Analysis

Based on the sample of 849 students, mean scores and standard deviations were obtained and test reliability for the total scores of four modes of the General Shame Scale (88 items) was calculated. The internal consistency reliability was estimated, using coefficient alpha (Cronbach, 1951). The results are shown in Table 11. In the analysis of variance, the F statistic for the variation between items was significant ($F = 152.86$, $p < 0.001$). This indicated that the items had significantly different means. This finding was confirmed by the large Hotelling's T-squared statistic ($T\text{-squared} = 4084.47$) which is a test for the equality of means. Its F statistic ($F = 42.19$, $p < 0.001$) was significant and indicated that the hypothesis that the items had equal means in the population could be rejected. The 88-item test was reliable with Cronbach's alpha at 0.9670.

Table 11. Results of Reliability Analysis on Four Modes of General Shame Scale (88 Items) in the Questionnaire

Analysis of Variance					
Source of Variation	Sum of Sq.	DF	Mean Square	F	Prob.
Between People	27606.2646	848	32.5546		
Within People	93563.2159	73863	1.2667		
Between Measures	14289.9870	87	164.2527	152.8626	.0000
Residual	79273.2289	73776	1.0745		
Total	121169.4806	74711	1.6218		
Grand Mean	3.3890				
Hotelling's T-Squared = 4084.4675 F = 42.1867 Prob. = .0000					
Degrees of Freedom:		Numerator = 87	Denominator = 762		
Reliability Coefficients 88 items					
Alpha = .9670	Standardized item alpha = .9682				

Another reliability analysis was conducted on each of the four different modes of the questionnaire, the anxiety factor, and the life satisfaction factor. The results are shown in Table 12. The Cronbach's alpha coefficients for the four modes were 0.9392 (Mode 1), 0.9059 (Mode 2), 0.9383 (Mode 3), and 0.8891 (Mode 4). This showed that the four modes had about the same reliability. The Cronbach's alpha coefficients for the anxiety factor and life satisfaction factor were 0.7524 and 0.6268 respectively.

Table 12. Mean, Standard Deviation, and Cronbach's Alpha Coefficients for the Subscales

Subscale	Mean	Standard Derivation	Cronbach's alpha coefficients
Mode 1	68.50	17.37	0.9392
Mode 2	78.70	14.84	0.9059
Mode 3	83.10	16.15	0.9383
Mode 4	67.93	15.80	0.8891
DOSC-Anxiety Factor	34.51	8.51	0.7524
Life Satisfaction	14.50	3.89	0.6268

Correlational Analyses

Items representing each mode were summed to obtained a subtotal score for each mode. Total test scores were obtained by adding the four subtotal scores. It was reasoned that if subjects' responses were mainly determined by item contents, then all four categories of items would tap the same source trait (construct), and, consequently, the correlation coefficients among the pairs of mode subtotal scores would be high; otherwise, they would be low. Mode 1 and Mode 4 variables were assumed to be positive, and Modes 2 and 3 were assumed to be negative aspects of the same construct.

Pearson correlation coefficients were obtained for each of these subscales (Mode 1 to Mode 4), DOSC-Anxiety Factor, Satisfaction with Life Scale, and gender. By inspecting these correlations, it became clear that the subscales correlated among each other. The result of the correlation analyses is presented in Table 13. The Pearson correlation coefficients ranged from -0.133 to 0.898. All subscales were significantly correlated with each other, except between Satisfaction with Life Factor and gender. Gender had low correlation coefficients with the four modes, DOSC, and Satisfaction with Life Factor, ranging from -0.133 to 0.128. There seems to be not much relationship between gender and these scales.

Table 13. Correlations between the 4 modes, DOSC-Anxiety Factor, Satisfaction with Life Scale, and Gender

	Mode 1	Mode 2	Mode 3	Mode 4	DOSC	Life	Gender
Mode 1	1.000						
Mode 2	0.660	1.000					
Mode 3	0.672	0.898	1.000				
Mode 4	0.656	0.315	0.346	1.000			
DOSC	0.325	0.465	0.491	0.166	1.000		
Life	0.616	0.452	0.442	0.345	0.219	1.000	
Gender	-0.110	-0.133	-0.129	-0.055	0.128	-0.002	1.000

The major problem this study set out to investigate concerned the construct validity of the differential semantic modes of item presentation. Among six correlation coefficients between the pairs of four semantic mode scores, all are significantly different

from zero ($p < 0.05$). However, the double-negative Mode 4 item score showed a smaller relationship with both Mode 2 and Mode 3 item scores. Mode 2 and Mode 3 contained, overall, negative-valence items, whereas Modes 1 and 4 had positive-valence items. The largest correlation coefficient in the set (0.898) was between Mode 2 and Mode 3. On the basis of correlational evidence, it is conspicuous that, by and large, Mode 4 seems to measure different constructs from Mode 2 and Mode 3 in spite of their deceptive content similarity.

On the other hand, the value (0.97) of coefficient alpha, which is generally taken as an evidence of the homogenous nature of the test components, is quite impressive. An inspection of the item-remainder correlation coefficients revealed that 8 of the 88 items have corrected item-total correlations of less than .35 (Appendix F). Out of these eight items, two belong to Mode 2 and six items are in Mode 4 (Table 14). The salient common feature of these modes is that they are characterized by the grammatically negative form. This produces evidence in favor of the argument that the “do not” form of sentences, even with a simple structure otherwise, creates ambiguity and confusion, and that double negatives, as in Mode 4, add to this confusion.

Table 14. Items with Corrected Item-Total Correlation Less Than 0.35.

Mode 2	Item	Corrected Item Total Correlation
	Q64. I do not feel significant enough to others that people notice me.	0.2249
	Q89. I don't see myself as intact and without personal defects.	0.1127
Mode 4	Q1. I do not see myself as flawed and with personal defects.	0.2599
	Q5. I am not unhappy with myself.	-0.4999
	Q6. I do not always fall short of my aspirations.	0.1564
	Q27. I do not feel unimportant.	0.3383
	Q56. Compared to others, I do not feel like I am less of a person.	0.2797
	Q71. I don't feel inferior to most of my friends.	0.3006

Research Question 2: What is the item equivalence between different modes of item wording?

MANOVA Analyses

One may hypothesize that if all four modes of presentation measured the same construct in the same way in all the groups of students, then one should be able to reach the same conclusions on the basis of any of the previously defined, five dependent variables (four mode subtotal scores and the total score). On the basis of male and female as the group determiners, a Multivariate Analysis of Variance procedure (MANOVA) was conducted. The results of the MANOVA are presented in Tables 15 and 16.

Inspection of each row in Table 16, showing F ratios calculated from four different sets of scores (four item modes), reveals that the four F-ratio values were far from being equivalent. The MANOVA results showed that the responses to the modes of item wording are significantly different between male and female. The F-ratios of Mode 1, Mode 2, and Mode 3 were all significant at the 0.05 level. Mode 4 was not significant at the 0.05 level. These results were similar to that of the correlational analyses. It seems that Mode 4, which has a double negative semantics, introduced some ambiguity to the items.

Table 15. Descriptive Statistics of Four Semantic Modes by Gender

	Gender	Mean	Std. Deviation
Mode 1	Male	70.3273	16.6181
	Female	66.5025	17.9645
	Total	68.4982	17.3704
Mode 2	Male	80.5553	14.1062
	Female	76.6010	15.3556
	Total	78.6643	14.8404
Mode 3	Male	85.1400	15.1455
	Female	80.9606	16.9337
	Total	83.1413	16.1517
Mode 4	Male	68.7585	16.0670
	Female	67.0222	15.4791
	Total	67.9282	15.8031

Table 16. MANOVA Tables by Gender for the Four Semantic Mode Subtotal Scores**Multivariate Tests**

Effect		Value	F	Hypothesis df	Error df	Sig.	Eta Squared
Intercept	Pillai's Trace	.975	8215.575	4.000	844.000	.000	.975
	Wilks' Lambda	.025	8215.575	4.000	844.000	.000	.975
	Hotelling's Trace	38.936	8215.575	4.000	844.000	.000	.975
	Roy's Largest Root	38.936	8215.575	4.000	844.000	.000	.975
Gender	Pillai's Trace	.019	4.062	4.000	844.000	.003	.019
	Wilks' Lambda	.981	4.062	4.000	844.000	.003	.019
	Hotelling's Trace	.019	4.062	4.000	844.000	.003	.019
	Hotelling's Trace	.019	4.062	4.000	844.000	.003	.019
	Roy's Largest Root	.019	4.062	4.000	844.000	.003	.019
					0		

Table 16 (Cont'd)**Tests of Between-Subjects Effects**

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	Mode 1	3099.210	1	3099.210	10.385	.001	.012
	Mode 2	3312.574	1	3312.574	15.294	.000	.018
	Mode 3	3700.347	1	3700.347	14.409	.000	.017
	Mode 4	638.661	1	638.661	2.562	.110	.003
Intercept	Mode 1	3966279.422	1	3966279.422	13290.652	.000	.940
	Mode 2	5232215.283	1	5232215.283	24157.626	.000	.966
	Mode 3	5844726.448	1	5844726.448	22758.468	.000	.964
	Mode 4	3905689.591	1	3905689.591	15667.897	.000	.949
Gender	Mode 1	3099.210	1	3099.210	10.385	.001	.012
	Mode 2	3312.574	1	3312.574	15.294	.000	.018
	Mode 3	3700.347	1	3700.347	14.409	.000	.017
	Mode 4	638.661	1	638.661	2.562	.110	.003
Error	Mode 1	252767.037	847	298.426			
	Mode 2	183448.755	847	216.586			
	Mode 3	217522.692	847	256.815			
	Mode 4	211139.956	847	249.280			
Total	Mode 1	4239381.000	849				
	Mode 2	5440436.000	849				
	Mode 3	6089921.000	849				
	Mode 4	4129263.000	849				
Corrected Total	Mode 1	255866.247	848				
	Mode 2	186761.329	848				
	Mode 3	221223.039	848				
	Mode 4	211778.617	848				

Exploratory Factor Analyses

The survey was subjected to a principal components analysis as the initial method of factor extraction. The two criteria used to select the model were obtained from suggestions offered by Rummel (1970) and Hair, Anderson, and Tatham (1987), that is, a) statistical indicators for selecting the 'best' number of factors, and b) an analysis of the content of the factors. Selection rules for the best number of factors have been developed for eigenvalues, scree plots, percent of variance accounted for by the model, overlap in factor loadings, and loadings values. The following selection rules applied. First, a factor must have an eigenvalue of one or greater to be considered significant. Second, a scree plot indicates the maximum number of factors to extract when the plot becomes horizontal, that is, when the curve first begins to straighten out. As a general rule, the scree tail test will result in at least one more factor being considered significant than will the latent root criterion (Hair et al., 1987). Third, the percent of variance accounted for should be as great as possible in considering the best number of factors. Fourth, the choice with the best number of factors should have the least amount of overlap in the total item factor loadings. Finally, the loading value with the largest absolute factor loading is the optimal choice. The loading values should be at least 0.30 to be considered significant, while factor loadings of 0.50 or greater are considered very significant. Ultimately, the number of significant loadings on each column of the factor matrix or loading associated with one variable would need to be maximized (Hair et al., 1987). Thus, a dually loaded item would be placed in the factor with the higher loading.

Product-moment correlation coefficients were computed, and a principal components analysis with iterations was performed on the resulting correlation matrix. The 107-item questionnaire was subjected to a principal components analysis. The factor analysis yielded nineteen factors with eigenvalue greater than 1.00. Three items from DOSC (question #11, #15, & #74) had no factor loadings over 0.30 and hence could not be identified with any of the factors. The scree plot, which is the plot of the total variance associate with each factor, is shown in Figure 6. The factor structure of the questionnaire can be found in Table 17.

Figure 6. Scree Plot from the Exploratory Factor Analysis (19 Factors)

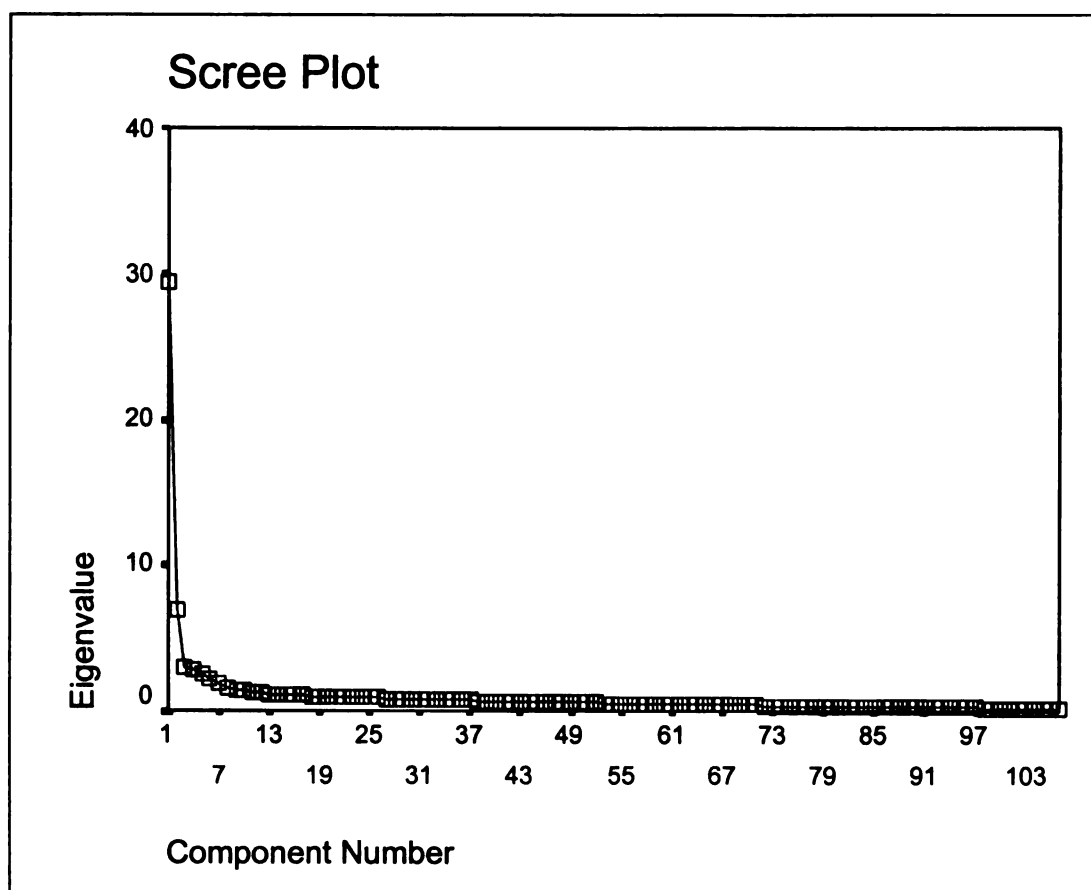


Table 17. Factor Structure of the Questionnaire (19 Factors)

Factor	Question Number
1	2,3,4,5,7,8,9,10,12,13,17,18,20,21,23,24,25,26,28, 30,31,32,33,34,35,36,37,39,41,42,44,46,49,50,51,52 54,55,60,61,66,68,72,73,75,79,87,88,92,93,94,97
2	1,14,16,27,40,47,53,56,63,71,91,98,99,100,101,104
3	57,58,59,102,107
4	38,62,65,69,70,89,90,105
5	43,48
6	71
7	64,85,103
8	76,84
9	80,96
10	67,95
11	6,106
12	83
13	86
14	82
15	78
16	45
17	22
18	81
19	77

Since the original factor structure was not interpretable, the number of factors was reduced to six in another analysis. This factor structure can be found in Table 18.

Table 18. Factor Structure of the Questionnaire (6 Factors)

Factor	Question Number
1	4,5,7,8,10,12,18,23,30,31,32,34,35,37,39,46,49,50, 51,52,54,55,57,59,60,61,66,72,75,77,78,82,86,92,93, 94,96,97,99,101
2	1,6,13,14,15,16,17,21,26,27,41,42,44,47,53,63,68,71,79,81,98,104,107
3	2,3,9,24,25,28,33,36,83,84,85,87,88,95
4	11,38,45,62,65,69,70,74,90,103,106
5	43,48,67,73,76,80,91,105
6	19,20,41,100,102

Confirmatory Factor Analyses

Confirmatory factor analysis may not be a complete answer to the issue of construct validity (See Hunter and Gerbing, 1982). The statistical method asks only whether items measure the same construct, not whether that construct is the right construct. Whether the items measure the right construct is a question of content, which is usually tested by looking at correlations between the scale and other constructs. At the level of the shame theme scales, the issue of the nature of the construct was dealt with solely in terms of an item content analysis. The items in each shame theme cluster were

closely examined to see if they were psychologically equivalent in both the affect expressed and in the manner in which that affect was expressed.

Hunter and Gerbing (1982) noted that if the right research design is used, then confirmatory factor analysis can be used to assess item equivalence in two very different ways: "internal consistency" and "parallelism" (or "external equivalence"). If item responses differ from each other only by random error of measurement, then the item errors will not correlate with each other. The correlations between items within a scale should then satisfy a mathematical product rule discovered by Spearman (1904, cited in Hunter & Gerbing, 1982): his one factor model. If the correlations between items within a scale satisfy the Spearman product rule, then the scale is said to be "internally consistent." However, this is a weak criterion for item equivalence.

There is a stronger criterion for item equivalence – parallelism in the pattern of correlations between the items and important "outside" variables, such as the measures of anxiety and life satisfaction factors used in the present study. If all items measure the same construct, then the item errors will not correlate with any outside variable. The correlations between items in a unidimensional scale with any outside variable should satisfy a condition called "parallelism" (Tryon, 1939; Tryon and Bailey, 1970; Hunter, 1973; Hunter and Gerbing, 1982). This is a strong test for item equivalence. If an item is contaminated by some unintended variable, and if that contaminating variable is one of the outside variables (or is correlated with one), then the item will correlate more highly with that outside variable than will the other uncontaminated items. Thus, failure to find parallelism not only shows an item to be contaminated, but it also identifies the contaminating variable (Hunter, 1986, 1987). Parallelism can either be tested by doing a

confirmatory factor analysis including both the items and the outside variables (as noted by Hunter and Gerbing, 1982) or by computing the correlations between the items and the outside variables and examining that correlation matrix for parallelism (as noted by Hunter, 1973, and Tryon and Bailey, 1970).

To answer the question of whether the four Modes of semantics measured the same construct, five models were tested. The first model specified a single factor in which all the items in four different modes measured one single general factor. The second model, specified a two-factor model where one factor represented the positive wording items and the other factor represented the negative wording items. The third model specified another two-factor model with one factor representing double negatives (Mode 4), and the second factor with affirmative semantic mode (Mode 1 and Mode 3) and with the “don’t” form semantics (Mode 2). The fourth model, specified three factors: one factor representing affirmative semantic mode (Mode 1 and Mode 3), the second with the “don’t” form semantics (Mode 2), and the third factor with double negatives (Mode 4). The fifth model parameterized the four Modes of semantics in the survey.

Results from these five sets of confirmatory factor analyses models are reported in Table 19. The 2-factor model (Modes 1, 2, & 3 vs. Mode 4) fit the data statistically and showed an overwhelming superiority over the other models. These results rendered strong indications of the inequivalence between double negatives (Mode 4) and the rest of the items (Modes 1, 2, & 3). The 3-factor model also fit the data statistically. The results rendered some indications of the inequivalence between the affirmative semantic mode (Mode 1 & 3), “don’t” form semantics (Mode 2), and double negatives (Mode 4).

Table 19. Goodness-of-Fit Indices of Five Models

Model	Chi-square	GFI	AGFI
1-factor (Modes 1,2,3,4 as one general factor))	298.30	0.721	0.699
2-factor (Modes 1 & 4 vs. Modes 2 & 3)	317.50	0.689	0.532
2-factor (Modes 1,2,3 vs. Mode 4)	222.67	0.956	0.889
3-factor (Modes 1 & 3, vs. Mode 2 vs. Mode 4)	259.70	0.938	0.874
4-factor (Modes 1,2,3,4 as individual factor)	389.76	0.679	0.514

Note. GFI=goodness-of-fit index; AGFI=adjusted goodness-of-fit test

Summary

The data of this study were the responses of 849 students to a questionnaire.

Preliminary steps were taken to ensure that the data were appropriately coded and sufficiently accurate before the start of the analysis. Statistical data analysis techniques of exploratory factor analysis, confirmatory factor analysis, analysis of variance, and reliability analysis were employed to answer the research questions.

CHAPTER V

SUMMARY, CONCLUSIONS, IMPLICATIONS, AND RECOMMENDATIONS

Summary of the Purposes and Procedures of the Study

A group of 849 Taiwanese students from two schools were selected for this study. These two schools were located in an urban setting. A questionnaire was constructed with 107 items. The students' responses to this 107-item questionnaire were analyzed. A confirmatory factor analysis was conducted and reliability of the subscales reported. Two research questions were formulated for this study. Statistical data analysis techniques of exploratory factor analysis, confirmatory factor analysis, analysis of variance, and reliability analysis were employed to answer the research questions.

The purpose of this study was to investigate how item wording affects rating responses. The first issue investigated was the influence of item wording on scale and item score mean, distribution, reliabilities, and correlations between scales. The second issue examined the item equivalence between different modes of item wording.

Discussion and Conclusion

Refinement of the survey development process is an essential component of any serious effort to enhance the reliability and validity of survey results. In the introductory chapter it was noted that item wording is an important consideration in survey development. A review of research on the impact of item wording on rating responses revealed a lack of consistent findings.

The results of this study suggest a rather important conclusion for measurement instrument design – that the inclusion of negatively worded items can result in less accurate responses and therefore can impair the validity of obtained results. Thus, although the inclusion of negatively stated items may theoretically control or offset agreement response tendencies, their actual effect is to reduce response validity. This situation suggests that current recommendations concerning the desirability of including both positive and negative items on a questionnaire may be premature (and perhaps incorrect), and the inclusion of both apparently warrants much further investigation.

In examining the effects of item wording on item responses, this study has shed some interesting light on an issue that has here-to-fore been the focus of arguments based upon ambiguous data and results. Overall, the findings suggest that the use of double negatively worded items may result in the measurement of a different construct than is intended. This outcome is in direct contrast to the conventional psychometric recommendations summarized earlier. This seems to be generally a function of the double negatively worded items themselves, and not of their exertion of a strong contextual effect on the positive items.

The data from the present study provide strong evidence that the insertion of the word “not” has a profound influence on the student responses. Two trends were indicated in the results. First, items that induced a more favorable response on the positive form induced a less favorable response on the negative form. In other words, respondents were less likely to indicate agreement by disagreeing with a negatively phrased item than to indicate agreement by agreeing with a positively phrased item. Second, items that induced an unfavorable response on the positive form were less likely to induce an unfavorable response on the negative form. In addition, confirmatory factor analyses indicated that the factor structures were clearly different for the positive and the negative forms.

The findings from the present study suggest that caution should be exercised in the use of negative item phrasing. Although it may be useful to include some negative items to reduce a response bias, these items need not be used in computing a total attitude score. The use of double negatively phrased items should be used with great caution. Despite the conventional wisdom so often found in measurement textbooks, recent pronouncements by researchers in the area of item phrasing have suggested that negatively phrased items, especially double negatives, reduce the validity of a questionnaire. This present study clearly corroborates that position.

The research to date suggests that positive to negative transformation changes an item’s psychometric characteristics, and, more importantly, changes the construct that an item is intended to measure. However, the present study and the other studies that have been reviewed do not prove that positively phrased items are necessarily better indicators of attitude. Nevertheless, there are some hints that negatively phrased items are less

valid. First, there is the plausible argument that respondents may not understand that they can indicate agreement by disagreeing with a negative statement. Similarly, they may not understand that they can indicate disagreement by agreeing with a negative statement.

A word of caution concerning the use of negatively-worded items is appropriate. If negatively worded items are to be used, it would be wise to ensure, during scale development, that their inclusion does not present a methodological confound. Rather than having alternative interpretations to contend with some two decades after scale development (e.g., McGee et al., 1989; Tracy & Johnson, 1981), it would be preferable to ensure that constructs are not exclusively defined by negatively worded items during scale development.

Implications

Survey development is a difficult task. This study attempts to provide some insights that will begin to answer questions pertinent to the impact of item wording on rating responses. The results of these analyses have a clear implication for researchers who factor analyze data in which the wording of items is varied. Such researchers should be cautious of factors loaded primarily with negatively keyed items. Likewise, consumers of this research should question substantive interpretations of such negative factors.

Researchers should be especially cautious concerning negative factors when responses to questionnaires are “involuntary” or when there is a reason to sabotage the researcher effort. The respondents’ data should be examined to detect unusual response patterns. If negative and positive items are recoded so as to be consistent, then a respondent whose primary responses on a 7-point scale are 5 and 6 would be suspicious if

negatively worded item responses were 2 and 3. Responses from these individual would be best deleted prior to any further analyses. A more systematic analysis of these respondents is possible with the use of item response theory. Latent trait analyses allow the determination of which item responses made by an individual are not well predicted by the IRT model. As a consequence, it is possible to detect unusual responses at the individual level. However, since ample size and number of item requirement for IRT analyses are large, latent trait parameters may not be obtainable for many instruments.

This investigation confirms the findings of earlier studies. Taken together, these studies offer important implications for measurement practice and theory development. One practical implication is that double negative items should not be used. Items on a test or survey should be consistent in positive wording with respect to the construct being measured. Positive and negative wording items are not bipolar indicators of a common trait continuum and therefore construct unidimensionality cannot be maintained by simply reversing the scale points associated with the negative items. Consequently, the inconsistent direction of wording is likely to change the intended factor structure of a test or survey. Researchers should not deliberately use double negatively worded items, even for the purpose of countering response-set effects.

Including many double negatively worded items in a test may have the impact of altering the original operational definition of the underlying construct. In order to maintain the intended original factor structure, the connotation of the items must be consistent, in one direction or the other. Research may be needed to reexamine the construct validity of other tests that use a large number of double negatively worded items.

It makes sense that the presence of a construct often indicated by a positively worded item does not necessarily mean the opposite (reversal of scale points) of the absence of a construct, as is often indicated by the negative wording counterparts on a test. It seems that the constructs in measurement are inseparable from the way their item indicators are connoted. Researchers using Likert-type scales routinely recode or flip the responses scales of the negatively worded items. Therefore, a primacy effect from this measurement factor needs to be avoided, considered, held constant, or separated from the interest effected by cautious researchers.

The results of this study will have important implications for researchers who analyze data in which the wording of items is varied. It is suggested that questionnaire instructions may include a warning to potential respondents that some questions will be negatively keyed and that they should attend to all items. The overall questionnaire length in an instrument that uses the same response format may also be a concern. The respondents may become fatigued or bored when they answer many like-sounding items. Trott and Jackson (1967) found that an acquiescence factor was strongly associated with the speed of presentation of personality items. It may be useful to experiment with the wording directions and the length of questionnaires/instruments as well as the serial position of any negatively keyed items. Further, the context in which data are collected could be varied in an effort to assess the effect of context on the presence or absence of negative factors.

The possible effect of item wording on overall ratings is particularly relevant to many of the student and employer rating instruments available today. Increasingly, emphasis is being placed upon the need for valid and reliable means of assessing teaching

and working performance. Rating scales used to evaluate a new project, person, or course of instruction, often include both negatively and positively stated items about the objects or person being evaluated. What has yet to be determined is the possible effect of item wording on raters' evaluations. Do negatively worded items encourage a more critical evaluation than do positively worded items? Negatively worded items may highlight the negative aspects or faults of the object or person being evaluated, or may serve to suggest unconsciously to the rater particular problem areas anticipated by the evaluator. If so, rating scale evaluations may be affected as much by the wording of the items as by the quality of the object or person being evaluated.

Several other researchers (Andrich, 1983; Campbell & Grissom, 1979; Simpson, Rentz, & Shrum, 1976) have investigated whether phrasing can influence overall attitude levels on different attitudinal questionnaires. These researchers have all concluded that item phrasing makes a difference. However, the results that these researchers report cannot be easily corroborated with each other or with the present study. One important differentiating feature is that in this study, the word "not" was used to create parallel negative statements, whereas the other researchers created negative statement on an intuitive basis. Rorer (1965) suggested that this latter procedure often leads to negative statements that reflect different content or ideas; consequently, such statements are not direct opposites of the original positive statement. It is perhaps because of this problem that many affective scales contain the word "not" or the prefix "un" to create a negative statement (Coopersmith, 1967; Marsh, Smith, Barnes & Butler, 1983; Piers, 1969).

Given the relative frequency with which a negative factor is reported in the literature and the ease with which a factor is produced, researchers should be especially

cautious when their factor analyses produce factors that are loaded primarily by negative items. Further, those who design questionnaires may also want to take steps to minimize problems during the construction of their instruments and with the directions which accompany them.

On the basis of the data analyses, the findings of this study conclude that double negative and semantically positive item contents do not measure essentially the same construct. Furthermore, these “do-not” form-generated negative-to-positive (double negative) items create ambiguity and confusion. This problem deserves closer inspection through more specifically designed studies. Interesting thoughts about measurement theory, suggested by the present study, may be worthy to ponder.

Cross Validation Using a New Sample

The problem of situational specificity is always a major concern in validity studies. Validity generalization research is based on the application of a particular set of meta-analytic methods (Hunter, Schmidt, & Jackson, 1982) to criterion-related validities of tests. This meta-analysis method was developed as a way of attacking a critically important problem in psychology: the problem of situation specificity. The belief was based on the empirical fact that considerable variability was present from study to study in observed validity coefficients, even when the jobs and tests studied appeared to be similar or identical. The explanation developed for this variability was that the factor structure of job performance was different from job to job. The conclusion was that validity studies must be conducted in every setting; that is, that validity evidence could not be generalized across settings.

Schmidt and Hunter (1981) hypothesized that most or all of the variance of study correlations across studies and settings was due to artifactual sources, such as sampling error, and not to real difference between jobs. Artifacts other than sampling error and differences between studies in measurement error and in range restriction can cause variance in study outcomes. Because the most common form of validity evidence is the correlation coefficient between predictor and criterion scores, it is important to recognize that the restriction of the range of scores on the questionnaire may results in attenuation of the observed validity coefficient. One example is the instance in which the test being validated is used for selection purposes before its validity has been established. Other things being equal, the greater the variability among the observations, the greater the value of correlation coefficients. Thus, restriction of range occurs on the questionnaire because of the explicit selection on that scale. As the correlations among items increase, the tests become more homogeneous in content (increase in internal consistency). Moreover, when we compute statistics from a set of data, we are getting the best estimates from the data. If these statistics are used in further calculations, we take advantage of the original calculations and therefore overestimate the second calculations. In Cureton's study (Cureton, 1950), the author said that one should not use a data set for conducting an item analysis and then use the results of that analysis to compute validity coefficients. When items are deleted from an original sample, some random errors are introduced. The cross validated sample will result in lower item correlations.

Cultural Specificity of the Questionnaire

Crocker and Algina (1986) mentioned that the ultimate criterion for the number of factors to interpret is replicability. When the same variables are investigated in different studies, the factors that are replicated in the studies are those that should be interpreted. Due to situational specificity of Taiwan, the survey may not generate the same factors shown by this study when applied to a new sample.

The items in the survey are very culturally specific. For instance, items such as "Compared to others, I feel like I don't measure up," "I find that I don't live up to my own standards and ideals," "I look down on myself because of my flawed character," "I feel so insignificant to others, as if I were invisible," and "I always seem to fall short of my aspirations," "I see myself as intact and without personal defects," while familiar to students of Chinese ethnicity, can have very different implications when administered to students of different cultures.

In general, Chinese think about social institutions such as school quite differently from American educators, seeing teachers as professionals with authority over their children's schooling. Teachers in the Chinese culture are accorded a higher status than teachers in the United States. They believe that parents are not supposed to interfere with school processes. Chinese people highly value formal education, and believe that high achievement brings honor and prestige to the family, while failure brings shame. The intense pressure upon children to succeed often leads to intergenerational conflicts, and many Chinese children suffer from test anxiety, social isolation, and low self-esteem because of their mediocre school performance. They have difficulties accepting learning

disabilities and depression, and believe that psychological distress is an indication of organic disorders and harmful to both the individual and the family.

Confucian ideals, which include respect for elders, deferred gratification, and discipline, are a strong influence. Cohesion and harmony are valued above individual achievement. Hard work, duty, obligation, frugality, and responsibility are also priorities. Most Chinese parents teach their children to value educational achievement, respect authority, feel responsibility for relatives, and show self-control. Chinese parents tend to view school failure as a lack of will, and they address this problem by increasing parental restrictions. Chinese children tend to be more dependent, conforming, and willing to place family welfare over individual wishes than are American children.

Self-effacement is a trait traditionally valued by Chinese culture. Chinese children tend to wait to participate, unless otherwise requested by the teacher. Having attention drawn to oneself, for example, having one's name put on the board for misbehaving, can bring considerable distress. Many Chinese children have been socialized to listen more than speak, to speak in a soft voice, and to be modest in dress and behavior. Discipline and obedience are highly valued in the Asian cultures, whereas creativity and freedom are important in Western cultures. The definition of appropriate student attitudes may be different due to differences in the interpretation of students' behaviors.

Issues Concerning the Use of Tests

Attitude measures are measures of typical behavior and are distinguished from the ability tests, which measure maximum performance (Cronbach, 1960). By measuring attitudes, we want to know what the person normally does rather than what he or she can do under exceptional motivation. Valid information about attitudes can be valuable to teachers, counselors, and students. Attitude inventories can help in identifying the problems and needs of students, provided that they are truthful in answering the items. These inventories provide a more complete and holistic understanding of the students. However, the results should not be treated as the sole source of information. Teachers and counselors can also identify problems and needs through observations and interviews. Since these tests usually have lower reliabilities and validities than cognitive tests, the interpretation of the results should be handled with great caution (Mehrens & Lehmann, 1984).

It is important for the technical quality of test materials and standards to be considered. The information about these should include evidence of reliability and validity; information regarding the method of estimating reliability and the population on which it was measured; and types of validity evidence, including validity relevant to the intended purpose of the test. Teachers and counselors need to ask themselves these questions: Is the test appropriate for the person who is being tested? How are the results going to be used? Are the test scores reliable enough? Does the test possess enough validity to be used for the purpose for which it is planned? Is the welfare of the student

being taken into consideration in the choice and use of tests? Will confidentiality become an issue if the subject does not want to reveal himself or herself to the tester?

Another important issue is the competence of the teacher or counselor who will be administering various available assessment instruments. Do those who use various tests have sufficient knowledge and understanding to select tests appropriately and to interpret their results? Since different tests demand different levels of competence for their use, users must recognize the limits of their competence and make use only of instruments for which they have adequate preparation and training.

Lastly, presentation of test results also requires significant attention. Teachers and counselors should avoid labeling when communicating the results of a test to students. Labeling can stigmatize a person even when such terms can be justified. They not only suggest a lack of any chance to grow or change, but they may also become self-fulfilling prophecies. Instead, interpretations should be presented in terms of possible ranges of academic achievement or formulations of interventions to assist the individual in behaving more effectively.

Limitations

As an investigation of the impact of item wording on rating responses, the present study has a number of limitations. First, since the survey was developed according to the context in the United States, the questionnaire may not be relevant to teachers' situations in other countries. Therefore, the application of this questionnaire to other cultures should be used with caution. Second, the subjects were students in Taiwan and they do not represent a random sample from the population. The results are therefore suggestive

rather than definitive, and cannot be generalized to other populations without qualification. Third, the factor structure generated by the factor analyses may represent a chance phenomenon, which would not hold up in a second study. A cross-validation study is needed in order to validate the generalizability of the factor structure. Fourthly, the conversion of item wording from one mode to another may add some confusion and ambiguity to the original meaning of the items. The translation of the questionnaire from English to Chinese may add more possible ambiguity to the original version of the questionnaire. Sometimes it is difficult to find an appropriate translation of certain words despite the best effort. Nonetheless, the results provide practical implications for test developers and measurement researchers.

Recommendations

The study has raised a number of issues which future work should address. This research sheds some light on the effects of item wording on rating responses, and suggests other possible investigations of problems of interest to survey developers and educational researchers. Several directions can be suggested for future research on the effects of item wording on rating responses. One direction would be to replicate the study using larger samples of students from countries other than Taiwan. Replication with other forms might shed more light on the pattern of interaction between each of these factors and form.

Only 107 items and 849 students were used in this study. A relevant question that might be asked is to what extent can the results of the study be generalized to students in other cultures. Thus another direction which future research might take would be to

replicate the study using subjects and items from other disciplines, such as politics and religion, which are more bipolar.

Further research might reveal answers to such questions as:

- 1) Will the factor structure remain the same when the questionnaire is administered to another sample in a different setting?
- 2) How will the content of the items affect the item wording, which in turn may affect the item responses?
- 3) Is there a relationship between item content, item wording, and rating scale?
- 4) What are the differences between male and female students in responding to this instrument?

Despite the work that still needs to be done in this area, this study provides some insights in the field of survey development. Researchers may be able to gain new insights resulting in more efficient survey construction. Further research into the item wording and the item responses would not only improve the accuracy of the inferences that may be made from these surveys, but also may have an important impact on survey development.

APPENDICES

APPENDIX A

Four Modes of Item Wording

Table 20. Four Modes of Item Wording

MODE 1	MODE 2	MODE 3	MODE 4
I like myself	I don't like myself	I dislike myself	I don't dislike myself
I am pleased with myself	I am not pleased with myself	I am unhappy with myself	I am not unhappy with myself
I am satisfied with myself	I am not satisfied with myself	I am disappointed with myself	I am not disappointed with myself
I feel proud of myself	I do not feel proud of myself	I feel ashamed of myself	I do not feel ashamed of myself
I feel that I am good enough	I do not feel I am good enough	I feel that I am just not good enough	I do not feel that I am just not good enough
I feel that I am superior to most of my friends	I do not feel that I am superior to most of my friends	I feel that I am inferior to most of my friends	I don't feel inferior to most of my friends
Compared to others, I feel like I measure up	Compared to others, I feel like I don't measure up	Compared to others, I feel like I am less of a person	Compared to others, I don't feel like I am less of a person
I am just as good as my friends	I am not as good as my friends	I am inferior to my friends	I am not inferior to my friends
I feel adequate as a person	I don't feel adequate as a person	I feel inadequate as a person	I don't feel I am inadequate as a person.
I feel that my character is intact	I don't feel my character is intact	I feel there is something defective in my character	I don't feel there is something defective in my character
I look up to myself because of my good character	I do not look up to myself because of my flawed character	I look down on myself because of my flawed character	I do not look down on myself because of my good character
I see myself as intact and without personal defects	I do not see myself as intact and without personal defects	I see myself as flawed and with personal defects	I do not see myself as flawed and with personal defects
I feel worthy as a person	I do not feel worthy as a person	I feel worthless as a person	I do not feel worthless as a person
I feel like a useful person	I do not feel like a useful person	I feel like a useless person	I do not feel like a useless person

Table 20. (cont'd)

MODE 1	MODE 2	MODE 3	MODE 4
I feel like I am good for something	I do not feel like I am good for something	I feel like I am good for nothing	I do not feel like I am good for nothing
I feel I am a complete success as a person	I do not feel I am a complete success as a person	I feel like I am a complete failure as a person	I do not feel like I am a complete failure as a person
I feel successful	I do not feel successful	I feel like a failure	I do not feel like a failure
I am a worthwhile person	I am not a worthwhile person	I am a worthless person	I am not a worthless person
I feel very important	I do not feel important	I feel unimportant	I do not feel unimportant
I feel so significant to others that people notice me	I do not feel significant enough to others that people notice me	I feel insignificant to others, as if I were invisible	I do not feel insignificant to others, as if I were invisible
I always seem to live up to my aspirations	I don't always live up to my aspirations	I always fall short of my aspirations	I don't always fall short of my aspirations
I find that I live up to my own standards and ideals	I find that I don't live up to my standards and ideals	I find that I fall short of my own standards and ideals	I find that I do not fall short of my own standards and ideals

APPENDIX B

Questionnaire in English

Code Number: _____

INSTRUCTIONS

For each item, fill in the circle on the answer sheet for that item which corresponds to the word or phrase that best describes yourself. Read the response options carefully before making your selection. These survey results will be used in a research study. Please read both the item and response options carefully before selecting your answer.

Your answers will be kept strictly confidential. Results of this survey will appear in summary or statistical form only, so that individuals cannot be identified. Thank you for your time and cooperation.

TO THE STUDENT

There are five possible responses to each statement:

Never	Seldom	Sometimes	Often	Almost Always
1	2	3	4	5
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

For each statement select ONE response. Please mark the bubble which best indicates your agreement with the statement. There are no right or wrong responses.

	Never 1	Seldom 2	Sometimes 3	Often 4	Almost Always 5
1. I do not see myself as flawed and with personal defects.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. I am pleased with myself.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. I feel that my character is intact.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. I am not as good as my friends.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. I am not unhappy with myself.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. I do not always fall short of my aspirations.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. I feel like a failure.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. I am a worthless person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. I am look up to myself because of my good character.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. I do not feel proud of myself.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11. I worry that my score on a test will not be one of the highest in class.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12. I feel there is a something defective in my character.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13. I am a worthwhile person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
14. I do not feel like I am good for nothing.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
15. Statements that some teachers make about my schoolwork hurt my feelings.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
16. I do not feel ashamed of myself.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
17. I am not disappointed with myself.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
18. I am inferior to my friends.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
19. Talking in front of class makes me feel nervous.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Never	Seldom	Sometimes	Often	Almost Always
	1	2	3	4	5
20. Compared to others, I feel like I don't measure up.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
21. I do not look down on myself because of my good character.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
22. If I could live my life over, I would change almost nothing.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
23. I do not feel important.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
24. I like myself.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
25. I feel I am a complete success as a person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
26. I do not feel like a failure.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
27. I do not feel unimportant.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
28. I am satisfied with my life.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
29. I am embarrassed to face my friends or family if I have made a low grade on a test or assignment.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
30. I feel unimportant.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
31. I don't feel adequate as a person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
32. I see myself as flawed and with personal defects.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
33. The conditions of my life are excellent.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
34. I don't feel like I am good for something.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
35. I feel inadequate as a person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
36. I am satisfied with myself.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
37. I feel that I am just not good enough.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
38. I have trouble sleeping well the night before an important examination.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
39. I feel like I am good for nothing.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Never	Seldom	Sometimes	Often	Almost Always
	1	2	3	4	5
40. I do not feel like a useless person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
41. Compared to other people I feel like I do measure up.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
42. I don't dislike my self.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
43. I find that I don't live up to my own standards or ideals.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
44. I do not feel I am a complete failure as a person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
45. I avoid talking to my classmates about schoolwork because they might make fun of me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
46. I am not satisfied with myself.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
47. I am not a worthless person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
48. I find that I live up to my own standards and ideals.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
49. I feel like I am good enough.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
50. I feel ashamed of myself.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
51. I do not like myself.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
52. I feel like a useless person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
53. I don't feel there is a something defective in my character.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
54. Compared to others, I feel like I am less of a person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
55. I feel so insignificant to others, as if I were invisible.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
56. Compared to others, I do not feel like I am less of a person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
57. I feel worthy as a person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
58. So far I have gotten the important thing I want in life.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Never	Seldom	Sometimes	Often	Almost Always
	1	2	3	4	5
59. I am not a worthwhile person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
60. I don't feel worthy as a person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
61. I feel like a useful person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
62. I feel so nervous about some of my classes that it is hard for me to attend.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
63. I do not feel that I am just not good enough.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
64. I do not feel significant enough to others that I people notice me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
65. I am afraid to ask teachers to explain a difficult concept a second or third time.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
66. I feel that I am inferior to most of my friends.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
67. I always seem to live up to what I aspire to be.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
68. I don't feel I am defective as a person, as if something is basically wrong with me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
69. I become frightened when a teacher calls on me in class.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
70. I am upset about so many things that I cannot concentrate on or do my schoolwork.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
71. I don't feel inferior to most of my friends.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
72. I don't feel I am a complete success as a person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
73. I find that I fall short of my own standards or ideals.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
74. I feel upset when I have to take a test.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
75. I dislike my self.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
76. In most ways my life is close to my ideal.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
77. I feel I am a complete failure as a person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Never	Seldom	Sometimes	Often	Almost Always
	1	2	3	4	5
78. I do not feel that I am superior to most of my friends.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
79. I do not feel worthless as a person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
80. I always seem to fall short of my aspirations.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
81. I see myself as intact and without personal defects.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
82. I don't feel like a useful person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
83. I feel proud of myself.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
84. I am unhappy with myself.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
85. I feel successful.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
86. I don't look up myself because of my flawed character flaws.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
87. I feel important.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
88. I feel adequate as a person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
89. I don't see myself as intact and without personal defects.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
90. I become tense and nervous when I am studying.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
91. I find that I don't fall short to my own standards or ideals.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
92. I look down on myself because of my flawed character.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
93. I do not feel that I am good enough.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
94. I feel disappointed with myself.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
95. I am just as good as my friends.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
96. I am not pleased with my self.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
97. I feel worthless as a person.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
98. I am not inferior to my friends.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Never	Seldom	Sometimes	Often	Almost Always
	1	2	3	4	5
99. I do not feel successful.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
100. I feel so significant that people notice me.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
101. I don't feel my character is intact.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
102. I feel I am superior to most of my friends.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
103. I worry about how well I am doing in my classes.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
104. I feel like I am good for something.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
105. I don't always seem to live up to my aspirations.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
106. I would be afraid to tell a teacher that he or she made a mistake in explaining an assignment or in working a problem.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
107. I don't feel insignificant to others, as if I were invisible.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Gender: _____

Grade: _____

Age: _____

APPENDIX C

Questionnaires in Chinese

編號：_____

作答指示

請閱讀每一題描述有關你的句子，然後將相應的圓圈塗滿。在選擇之前，請仔細閱讀每一個選項。這份問卷的結果將會作研究之用。在你作答之前，請仔細閱讀每一個句子與選項。

你的答案將會絕對保密。問卷結果將以摘要或統計方式呈現。所以個人答案不會被列出。謝謝你的參與和合作。

給同學的指示

每一題都有五個可能的選項：

從不	偶然	有時	通常	常常
1	2	3	4	5

每一題請選擇一個選項。請將你認為最適合的相應選項之圓圈塗滿。所有選擇都沒有對錯之分。

	從不 1	偶然 2	有時 3	通常 4	常常 5
1. 我不看自己是無能而有缺陷的人	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. 我對自己高興	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. 我覺得自己人格完整	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. 我沒有我的朋友那麼好	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. 我對自己不開心	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. 我不是常常缺乏抱負	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. 我感覺失敗	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. 我是一個沒有價值的人	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. 我因自己的完整人格看得起自己	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. 我不以自己為榮	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11. 我擔心我的考試成績不是全班最好中的一個	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
12. 我覺得自己人格有缺陷	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
13. 我是一個有價值的人	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
14. 我不覺得自己一事無成	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
15. 某些老師對我學校功課的評語傷害了我	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
16. 我不自己為恥	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
17. 我對自己不失望	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
18. 我比不上我的朋友	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
19. 在全班同學面前講話令我緊張	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	從不 1	偶然 2	有時 3	通常 4	常常 5
20. 與他人比較我不覺得自己 高人一等	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
21. 我不因自己的不完整人格 看不起自己	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
22. 直到目前為止我都不能得 到生命中認為重要的東西	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
23. 我不覺得自己很重要	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
24. 我喜歡自己	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
25. 我覺得自己很成功	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
26. 我不感覺失敗	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
27. 我不覺得自己不重要	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
28. 我對我的生活感覺滿意	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
29. 如果考試或功課成績不好 我會覺得很慚愧去面對 朋友或家人	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
30. 我覺得自己不重要	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
31. 我不覺得自己有足夠能力	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
32. 我看自己是無能而有缺陷 的人	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
33. 我的生活條件極佳	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
34. 我不覺得自己可以有所成就	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
35. 我覺得自己無能	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
36. 我對自己滿意	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
37. 我覺得自己沒有能力	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
38. 每當有重要考試時我前 一晚都睡不好	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
39. 我覺得自己一事無成	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	從不 1	偶然 2	有時 3	通常 4	常常 5
40. 我不覺得自己是無用的人	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
41. 與他人比較我覺得自己 高人一等	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
42. 我不討厭自己	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
43. 我發覺我不為自己的原則與 理想而活	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
44. 我不覺得自己很失敗	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
45. 我避免與同學討論功課因為 怕他們取笑我	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
46. 我對自己不滿意	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
47. 我不是一個沒有價值的人	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
48. 我發覺我為自己的原則與 理想而活	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
49. 我覺得自己很好	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
50. 我以自己為恥	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
51. 我不喜歡自己	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
52. 我覺得自己是無用的人	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
53. 我不覺得自己人格有缺陷	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
54. 與他人比較我覺得自己比不 上人家	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
55. 我覺得自己不重要像不 存在一樣	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
56. 與他人比較我不覺得自己比 不上人家	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
57. 我覺得自己有價值	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
58. 直到目前為止我都能得到 生命中認為重要的東西	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	從不 1	偶然 2	有時 3	通常 4	常常 5
59. 我不是一個有價值的人	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
60. 我不覺得自己有價值	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
61. 我覺得自己是有用的人	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
62. 我對某些科目過於緊張以至無法上課	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
63. 我不覺得自己沒有能力	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
64. 我不覺得自己重要到令人們注意我	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
65. 我不敢請老師解釋一個困難的問題第二或第三遍	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
66. 我覺得自己比不上很多朋友	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
67. 我常常為抱負而活	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
68. 我不覺得自己無能	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
69. 當老師在班上叫我時我會很害怕	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
70. 很多事令我生氣以至我無法專心	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
71. 我不覺得自己比不上很多朋友	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
72. 我不覺得自己很成功	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
73. 我發覺我缺乏自己的原則與理想	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
74. 當我要考試時我會覺得不舒服	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
75. 我討厭自己	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
76. 在很多方面我的生活已接近理想	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
77. 我覺得自己很失敗	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	從不 1	偶然 2	有時 3	通常 4	常常 5
78. 我不覺得自己好過很多朋友	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
79. 我不覺得自己沒有價值	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
80. 我常常缺乏抱負	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
81. 我看自己是有能力而沒有缺陷的人	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
82. 我不覺得自己是有用的人	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
83. 我以自己為榮	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
84. 我對自己開心	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
85. 我感覺成功	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
86. 我因自己的不完整人格 不看得起自己	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
87. 我覺得自己很重要	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
88. 我覺得自己有足夠能力	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
89. 我不看自己是有能力而沒有缺陷的人	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
90. 在讀書時我會變得異常緊張	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
91. 我發覺我並不缺乏自己的 原則與理想	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
92. 我因自己的不完整人格 看不起自己	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
93. 我不覺得自己很好	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
94. 我對自己失望	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
95. 我像我的朋友一樣好	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
96. 我對自己不高興	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
97. 我覺得自己沒有價值	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
98. 我不會比不上我的朋友	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	從不 1	偶然 2	有時 3	通常 4	常常 5
99. 我不感覺成功	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
100. 我覺得自己很重要以至 人們注意我	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
101. 我不覺得自己人格完整	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
102. 我覺得自己好過很多朋友	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
103. 我擔心自己在功課方面 做得不夠好	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
104. 我覺得自己可以有所成就	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
105. 我不是常常為抱負而活	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
106. 當老師在解釋功課或例題 犯錯誤時我不敢告訴他	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
107. 我不覺得自己不重要像不 存在一樣	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

性別: _____

年級: _____

年齡: _____

APPENDIX D

Descriptive Statistics of the Items

Table 21. Descriptive Statistics of the Items

Item	Mean	SD	Variance
1	3.0212	1.4182	2.011
2	3.4511	1.1182	1.250
3	3.2768	1.1886	1.413
4	3.2556	1.0813	1.169
5	2.3934	1.0294	1.060
6	2.8363	1.2268	1.505
7	3.5053	1.0496	1.102
8	4.1567	1.0795	1.165
9	3.3934	1.2154	1.477
10	3.5995	1.1492	1.321
11	3.4087	1.4918	2.225
12	3.7538	1.1136	1.240
13	3.4570	1.2231	1.496
14	3.1390	1.3905	1.933
15	3.9199	1.1516	1.326
16	3.3816	1.3908	1.934
17	3.3675	1.1902	1.417
18	3.4134	1.1502	1.323
19	2.9870	1.4303	2.046
20	2.8728	1.2364	1.529
21	3.4912	1.3238	1.753
22	2.3840	1.4146	2.001
23	3.7503	1.2519	1.567
24	3.5783	1.2291	1.511
25	2.7491	1.0379	1.077
26	2.9411	1.1228	1.261
27	2.9835	1.3675	1.870
28	3.3922	1.1927	1.423
29	2.5689	1.2969	1.682
30	3.8799	1.1845	1.403
31	3.4064	1.1507	1.324
32	4.1519	1.0160	1.032
33	3.0777	1.1710	1.371
34	3.7986	1.1759	1.383

Table 21. (Cont'd)

Item	Mean	SD	Variance
35	4.2108	1.0252	1.051
36	3.1955	1.1470	1.315
37	3.9305	1.0511	1.105
38	3.9576	1.2046	1.451
39	3.9847	1.1121	1.237
40	3.2509	1.3879	1.926
41	2.3557	1.1173	1.248
42	3.5406	1.2935	1.673
43	3.6231	1.2021	1.445
44	3.1178	1.2411	1.540
45	4.2968	1.0054	1.011
46	3.6019	1.1642	1.355
47	3.2391	1.4027	1.968
48	3.2827	1.2683	1.609
49	3.1449	1.1570	1.339
50	4.1343	.9826	.965
51	4.0459	1.0078	1.016
52	4.1543	.9993	.999
53	3.0495	1.3484	1.818
54	3.4064	1.1352	1.289
55	3.8940	1.2279	1.508
56	2.9069	1.2226	1.495
57	3.4346	1.2212	1.491
58	2.9435	1.2350	1.525
59	3.9741	1.1261	1.268
60	3.9093	1.1574	1.340
61	3.4676	1.2050	1.452
62	4.1837	1.1208	1.256
63	3.1637	1.3685	1.873
64	3.1225	1.2790	1.636
65	3.0989	1.4801	2.191
66	3.4087	1.2137	1.473
67	2.9317	1.2785	1.634
68	3.3027	1.3835	1.914
69	3.4770	1.2121	1.469
70	3.0907	1.2738	1.623
71	2.9918	1.2853	1.652

Table 21. (Cont'd)

Item	Mean	SD	Variance
72	3.4287	1.1368	1.292
73	3.3899	1.2053	1.453
74	3.8587	1.1436	1.308
75	4.0342	1.0914	1.191
76	2.7044	1.1036	1.218
77	3.9258	1.0958	1.201
78	3.3663	1.1641	1.355
79	3.0224	1.3867	1.923
80	3.5701	1.1886	1.413
81	2.7409	1.3126	1.723
82	3.7915	1.1676	1.363
83	3.1390	1.2110	1.467
84	2.4935	1.1473	1.316
85	3.0389	1.0934	1.195
86	3.7703	1.1504	1.323
87	3.2874	1.2102	1.464
88	3.1602	1.1504	1.323
89	3.4158	1.2477	1.557
90	4.2686	1.0306	1.062
91	3.0707	1.2756	1.627
92	4.0047	1.1451	1.311
93	3.4747	1.1583	1.342
94	3.7244	1.1230	1.261
95	2.9882	1.2102	1.464
96	3.7880	1.0278	1.056
97	4.0141	1.0858	1.179
98	2.9788	1.2159	1.478
99	3.5807	1.1131	1.239
100	2.3581	1.1519	1.327
101	3.6219	1.1963	1.431
102	2.6396	1.2023	1.445
103	2.6631	1.2893	1.662
104	3.4276	1.2275	1.507
105	3.4664	1.2186	1.485
106	3.7126	1.2381	1.533
107	2.7385	1.4706	2.163

APPENDIX E

Descriptive Statistics of Four Modes of Item Wording

Table 22. Descriptive Statistics of Four Modes of Item Wording

<u>Items (Mode 1)</u>	Mode 1		Mode 2		Mode 3		Mode 4	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
I like myself.	3.5783	1.2291	4.0459	1.0078	4.0342	1.0914	3.5406	1.2935
I am pleased with myself.	3.4511	1.1182	3.7880	1.0278	2.4935	1.1473	2.3934	1.0294
I am satisfied with myself.	3.1955	1.1470	3.6019	1.1642	3.7244	1.1230	3.3675	1.1902
I fell proud of myself.	3.1390	1.2110	3.5995	1.1492	4.1343	.9826	3.3816	1.3908
I feel that I am good enough.	3.1449	1.1570	3.4747	1.1583	3.9305	1.0511	3.1637	1.3685
I feel that I am superior to most of my friends.	2.6396	1.2023	3.3663	1.1641	3.4087	1.2137	2.9918	1.2853
Compared to others, I feel like I measure up.	2.3557	1.1173	2.8728	1.2364	3.4064	1.1352	2.9069	1.2226
I am just as good as my friends.	2.9882	1.2102	3.2556	1.0813	3.4134	1.1502	2.9788	1.2159
I feel adequate as a person.	3.1602	1.1504	3.4064	1.1507	4.2108	1.0252	3.3027	1.3835
I feel that my character is intact.	3.2768	1.1886	3.6219	1.1963	3.7538	1.1136	3.0495	1.3484
I look up to myself because of my good character.	3.3934	1.2154	3.7703	1.1504	4.0047	1.1451	2.8728	1.2364
I see myself as adequate and without personal defects.	2.7409	1.3126	3.4158	1.2477	4.1519	1.0160	3.0212	1.4182
I feel worthy as a person.	3.4346	1.2212	3.9093	1.1574	4.0141	1.0858	3.0224	1.3867
I feel like a useful person.	3.4676	1.2050	3.7915	1.1676	4.1543	.9993	3.2509	1.3879

Table 22 (cont'd)

<u>Items</u>	Mode 1		Mode 2		Mode 3		Mode 4	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
I feel like I am good for something.	3.4276	1.2275	3.7986	1.1759	3.9847	1.1121	3.1390	1.3905
I feel I am a complete success as a person.	2.7491	1.0379	3.4287	1.1368	3.9258	1.0958	3.1178	1.2411
I feel successful.	3.0389	1.0934	3.5807	1.1131	3.5053	1.0496	2.9411	1.1228
I am a worthwhile person.	3.4570	1.2231	3.9741	1.1261	4.1567	1.0795	3.2391	1.4027
I feel very important.	3.2874	1.2102	3.7503	1.2519	3.8799	1.1845	2.9835	1.3675
I feel so important that people notice me.	2.3581	1.1519	3.1225	1.2790	3.8940	1.2279	2.7385	1.4706
I always seem to live up to my aspirations.	2.9317	1.2785	3.4664	1.2186	3.5701	1.1886	2.8363	1.2268
I find that I live up to my own standards and ideals.	3.2827	1.2683	3.6231	1.2021	3.3899	1.2053	3.0707	1.2756

APPENDIX F

Item Total Statistics

Table 23. Item-total Statistics (88 items)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
Q1	295.2108	2823.6147	.2599	.3038	.9672
Q2	294.7809	2803.9142	.5036	.5052	.9666
Q3	294.9552	2800.3754	.5009	.5405	.9666
Q4	294.9764	2816.4027	.4115	.4525	.9668
Q5	295.8386	2919.3454	-.4999	.4655	.9683
Q6	295.3958	2842.8314	.1564	.2749	.9673
Q7	294.7267	2795.8663	.6111	.5641	.9664
Q8	294.0754	2789.2820	.6521	.6412	.9663
Q9	294.8386	2786.4232	.5993	.5053	.9664
Q10	294.6325	2801.1077	.5128	.3768	.9666
Q12	294.4782	2809.2286	.4603	.5335	.9667
Q13	294.7750	2777.1156	.6686	.6023	.9663
Q14	295.0931	2810.9123	.3524	.3551	.9670
Q16	294.8504	2804.8915	.3935	.5178	.9669
Q17	294.8645	2799.2045	.5096	.4818	.9666
Q18	294.8186	2801.1345	.5121	.5478	.9666
Q20	295.3592	2815.3201	.3655	.4107	.9669
Q21	294.7409	2794.2653	.4915	.4781	.9666
Q23	294.4817	2783.4976	.6036	.5287	.9664
Q24	294.6537	2786.8115	.5894	.5456	.9664
Q25	295.4829	2794.5896	.6300	.5538	.9664
Q26	295.2909	2798.7938	.5450	.4635	.9665
Q27	295.2485	2813.8568	.3383	.3276	.9670
Q30	294.3522	2779.0350	.6755	.7050	.9663
Q31	294.8257	2802.9601	.4967	.4201	.9666
Q32	294.0801	2797.2412	.6190	.6095	.9664
Q34	294.4335	2794.1374	.5573	.4527	.9665
Q35	294.0212	2789.9312	.6816	.6864	.9663
Q36	295.0365	2779.5494	.6940	.6493	.9662
Q37	294.3015	2788.9609	.6732	.6447	.9663
Q39	294.2473	2786.0213	.6605	.6328	.9663
Q40	294.9812	2791.0963	.4894	.5006	.9666
Q41	295.8763	2802.5472	.5157	.4958	.9666
Q42	294.6914	2787.1264	.5565	.5248	.9665
Q43	294.6090	2816.8658	.3643	.4875	.9669
Q44	295.1143	2795.3985	.5171	.4718	.9666
Q46	294.6302	2794.6178	.5591	.5455	.9665
Q47	294.9929	2798.7028	.4321	.4327	.9668
Q48	294.9494	2800.0128	.4707	.5382	.9667
Q49	295.0872	2780.5797	.6793	.6238	.9663
Q50	294.0978	2809.2110	.5244	.4524	.9666
Q51	294.1861	2794.6941	.6484	.6698	.9664
Q52	294.0777	2791.3076	.6866	.7079	.9663
Q53	295.1826	2805.9065	.3995	.4268	.9668
Q54	294.8257	2788.2455	.6278	.6223	.9664
Q55	294.3380	2776.2240	.6729	.6598	.9663
Q56	295.3251	2826.9414	.2797	.3559	.9671
Q57	294.7974	2765.1146	.7646	.7344	.9661
Q59	294.2580	2790.1280	.6170	.5605	.9664
Q60	294.3227	2785.9193	.6346	.5532	.9663
Q61	294.7644	2772.1968	.7183	.6596	.9662
Q63	295.0683	2801.7383	.4224	.4447	.9668
Q64	295.1095	2832.5458	.2249	.3280	.9672
Q66	294.8233	2779.0371	.6587	.6387	.9663

Table 23. Item-total Statistics (88 items)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item- Total Correlation	Squared Multiple Correlation	Alpha if Item Deleted
Q67	295.3004	2810.9934	.3849	.4718	.9669
Q68	294.9293	2790.0658	.4982	.5109	.9666
Q71	295.2403	2822.0955	.3006	.3952	.9670
Q72	294.8033	2802.8327	.5041	.4995	.9666
Q73	294.8422	2802.7345	.4750	.5388	.9667
Q75	294.1979	2785.8405	.6751	.7021	.9663
Q77	294.3062	2782.5854	.7008	.7019	.9662
Q78	294.8657	2809.8640	.4342	.4262	.9667
Q79	295.2097	2808.0055	.3734	.4192	.9669
Q80	294.6620	2805.5118	.4596	.4987	.9667
Q81	295.4912	2804.8328	.4189	.3848	.9668
Q82	294.4405	2803.6005	.4839	.4000	.9666
Q83	295.0931	2776.9170	.6771	.6732	.9662
Q84	295.7385	2941.9410	-.6304	.6423	.9687
Q85	295.1932	2784.8164	.6828	.6663	.9663
Q86	294.4617	2790.1592	.6033	.5332	.9664
Q87	294.9446	2777.4014	.6737	.6537	.9663
Q88	295.0718	2785.3592	.6433	.5982	.9663
Q89	294.8163	2848.2303	.1127	.2130	.9674
Q91	295.1614	2800.9539	.4608	.4844	.9667
Q92	294.2273	2779.7584	.6934	.6727	.9662
Q93	294.7574	2791.0212	.5919	.5640	.9664
Q94	294.5077	2785.5403	.6580	.6557	.9663
Q95	295.2438	2789.8520	.5748	.4851	.9665
Q96	294.4441	2800.2472	.5837	.6450	.9665
Q97	294.2179	2782.3758	.7093	.7365	.9662
Q98	295.2532	2814.3474	.3796	.3736	.9669
Q99	294.6514	2796.5788	.5690	.5383	.9665
Q100	295.8740	2810.1457	.4367	.4556	.9667
Q101	294.6101	2805.5282	.4564	.4411	.9667
Q102	295.5925	2804.0979	.4654	.4837	.9667
Q104	294.8045	2782.4641	.6242	.5359	.9664
Q105	294.7656	2819.9650	.3350	.4455	.9669
Q107	295.4935	2829.4861	.2119	.2896	.9673

BIBLIOGRAPHY

BIBLIOGRAPHY

- Adorno, T. W., Frankel-Brunswik, E., Levinson, D. J., & Sanford, R. W. (1950). The Authoritarian Personality. New York: Harper.
- Ahlawat, K. S. (1985). On the negative valence items in self-report measures. Journal of General Psychology, 112(1), 89-99.
- Anastasi, A. (1982). Psychological Testing (5th ed.). New York: Macmillan.
- Anderson, L. W. (1981). Affective characteristics in the schools. Boston: Allyn & Bacon.
- Andrich, D. (1983). Diagnosing and accounting for response sets provoked by items of a questionnaire. Paper presented at the annual meeting of the National Council on Measurement in Education, Montreal. PQ.
- Andrulis, R. S. (1977). Adult assessment: A sourcebook of tests and measurement for human behavior. Springfield, IL: Thomas.
- Benson, J., & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary school children. Journal of Educational Measurement, 22, 213-240.
- Bentler, P. M., Jackson, D. N., & Messick, S. (1971). Identification of content and style: A two-dimensional interpretation of acquiescence. Psychological Bulletin, 76, 186-204.
- Bentler, P. M., Jackson, D. N., & Messick, S. (1972). A rose by any other name. Psychological Bulletin, 77, 109-113.
- Berg, I. A. (1961). Measuring deviant behavior by means of deviant response sets. In I. A. Berg & B. M. Bass (Eds.), Conformity and deviation (pp.328-397). New York: Harper and Row.
- Block, J. (1967). Remarks on Jackson's "review" of Block's challenge of response sets. Educational and Psychological Measurement, 27, 499-502.
- Block, J. (1971). On further conjecture regarding acquiescence. Psychological Bulletin, 76, 205-210.

- Block, J. (1972). The shifting definition of acquiescence. Psychological Bulletin, 78, 10-12.
- Bloom, B. S. (1978). New learner: Implications for instructions and curriculum. Educational Leadership, 35, 563-576.
- Campbell, N. O., & Grissom, S. (April, 1979). Influence of item direction on student responses in attitude assessment. Paper presented at the 63rd annual meeting of the American Educational Research Association, San Francisco, CA.
- Chang, L. (1995). Connotatively inconsistent test items. Applied Measurement in Education, 8(3), 199-209.
- Chang, S. S. and Hunter, J. (1988). Phenomenology and the measurement of shame. Unpublished manuscript.
- Coopersmith, S. (1967). The antecedents of self-esteem. San Francisco: Freeman.
- Couch, A., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. Journal of Abnormal and Social Psychology, 60, 151-174.
- Crocker, L. & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart, and Winston.
- Cronbach, L. J. (1946). Response sets and test validity. Educational and Psychological Measurement, 6, 475-494.
- Cronbach, L. J. (1950). Further evidence on response sets and test design. Educational and Psychological Measurement, 10, 3-31.
- Cronbach, L. J. (1960). Essentials of Psychological Testing (2nd. ed.). New York: Harper & Row.
- Danis, S. G. (1974). The effect of attitude and scale format on polarization in social judgments. Dissertation. University of Georgia.
- Diener, E., Emmons, R. A., Larson, R. J., & Griffin, S. (1985). The satisfaction with life scale. Journal of Personality Assessment, 49(1), 71-75.
- Dudycha, A. L., & Carpenter, J. B. (1973). Effects of item format on item discrimination and difficulty. Journal of Applied Psychology, 58, 116-121.

- Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. Journal of Applied Psychology, 37, 90-93.
- Edwards, A. L. (1955). Social desirability and Q-sorts. Journal of Consulting Psychology, 19, 464.
- Edwards, A. L. (1957a). The social desirability variable in personality assessment and research. New York: Dryden.
- Edwards, A. L. (1957b). Techniques of attitudes scale construction. New York: Appleton Century Crofts.
- Green, R. F. (1951). Does a selection situation induce testees to bias their answers on interest and temperament tests? Educational and Psychological Measurement, 11, 503-515.
- Guilford, J. P. (1936). Psychometric methods. New York: McGraw-Hill.
- Harasym, P. H. (1992). Evaluation of negation in stems of multiple-choice items. Evaluation and the Health Professions, 15(2), 198-220.
- Hathaway, S. R. and McKinley, J. C. (1967). MMPI Manual (Rev. ed). New York: Psychological Corporation.
- Hunter, J. E. (1973). Methods of reordering the correlation matrix to facilitate visual inspection and preliminary cluster analysis. Journal of Educational Measurement, 10, 51-61.
- Hunter, J. E., & Gerbing, D. W. (1982). Unidimensional measurement, second order factor analysis and causal models. In B. W. Saw, & L. L. Cummings (Eds.), Research in Organizational Behavior, Vol. 4. Greenwich, CN: JAI Press.
- Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). Meta-analysis: Cumulating research findings across studies. Beverly Hills, CA: Sage.
- Huttenlocher, J. (1962). Some effects of negative instances on the formation of simple concepts. Psychological Reports, 11, 35-42.
- Jackson, D. N. (1967a). Balance scales, item overlap and the stables of Auegas. Educational and Psychological Measurement, 27, 502-507.

- Jackson, D. N. (1967b). Block's challenge of response sets. Educational and Psychological Measurement, 27, 207-219.
- Jackson, D. N., & Messick, S. (1958). Content and style in personality assessment. Psychological Bulletin, 55, 243-252.
- Jackson, D. N., & Messick, S. (1965). The nonvanishing variance component. American Psychologist, 20, 498.
- Jackson, D. N., & Paunonen, S. V. (1980). Personality structure and assessment. Annual Review of Psychology, 31, 503-551.
- Jacobs, A., & Barron, R. (1968). Falsification of the Guilford-Zimmerman Temperament Survey: II. Making a poor impression. Psychological Reports, 23, 1271-1277.
- Jaroslogy, R. (1988. July/August). What's on your mind, America? Psychology Today, 54-59.
- Joreskog, K. G., & Sorbom, D. (1988). LISREL 7: A guide to the program and application. Chicago: SPSS Inc.
- Lemon, N. (1973). Attitudes and their Measurement. New York: John Wiley & Sons.
- Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 140, 152.
- Linn, R. L. (1990). Essentials of student assessment: From accountability to instructional aid. Teachers College Record, 91 (3), 422-436.
- Marsh, H. (1984). The bias of negatively worded items in rating scales for young children. Journal of Educational Psychology, 76, 420-431.
- Marsh, H., Smith, I., Barnes, J., & Butler, S. (1983). Self-concept: Reliability, dimensionality, validity and the measurement of change. Journal of Educational Psychology, 75, 772-790.
- Mehrens, W. A., & Lehmann, I. J. (1984). Measurement and Evaluation in Education and Psychology (3rd ed.). Orlando, FL: Holt, Rinehart, & Winston.
- Michael, W. B. and Smith R. A. (1976). The development and preliminary validation of three forms of a self-concept measure emphasizing school-related activities. Educational and Psychological Measurement, 36, 527-535.

- Michael, W. B., Denny, B., Ireland-Galman, M., & Michael, J. J. (1987). The factorial validity of a college-level form of an academic self-concept scale. Educational Research Quarterly, 11(1), 34-39.
- Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.
- Ory, J. C. (1982). Item placement and wording effects on overall ratings. Educational and Psychological Measurement, 42, 767-775.
- Osgood, C. E. (1952). The nature and measurement of meaning. Psychological Bulletin, 49, 197-237.
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). The measurement of meaning. Urbana: University of Illinois Press.
- Piers, E. (1969). The Piers-Harris children's self-concept scale. Nashville, TN: Counselor Recordings & Test, Box 6184, Acklen Station.
- Radcliffe, J. A. (1966). A note on questionnaire faking with 16 PFQ and MPI. Australian Journal of Psychology, 18, 154-157.
- Ramsay, J. O. (1973). The effect of number of categories in rating scale on precision of estimation of scale values. Psychometrika, 38, 513-529.
- Remmers, H. H., & Ewart, E. (1941). Reliability of multiple-choice measuring instruments as a function of the Spearman-Brown prophecy formula, III. Journal of Educational Psychology, 32, 61-66.
- Robinson, J. P., & Shaver, P. R. (1973). Measurement of social psychological attitudes. Ann Arbor, MI: Survey Research Center, Institute for Social Research.
- Rorer, L. G. (1965). The great response-style myth. Psychological Bulletin, 63, 129-156.
- Rotter, G. S. (1972). Attitudinal points of agreement and disagreement. Journal of Social Psychology, 86, 211-218.
- Rotter, G. S., & Barton, P. (1970). See attitudes of some New Jersey teachers. N.J.E.H. Review, 28-29.
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs, 80, 1.

- Samelson, F. (1972). Response style: A psychologist's fallacy. Psychological Bulletin, 78,13-16.
- Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. American Psychologist, 36, 1128-1137.
- Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The results of careless respondents? Applied Psychological Measurement, 9, 367-373.
- Schriesheim, C. A., & Hill, K. D. (1981). Controlling acquiescence response bias by item reversal: The effect on questionnaire validity. Educational and Psychological Measurement, 41, 1101-1114.
- Schriesheim, C. A. and Kerr, S. (1974). Psychometric properties of the Ohio State leadership scales. Psychological Bulletin, 81, 756-765.
- Scott, W. A. Attitude measurement. (1968) In G. Lindzey (Ed.), The handbook of social psychology, Vol.2 (2nd ed.). Reading, MA: Addison-Wesley.
- Shaw, M. E., & Wright, J. M. (1967). Scales for the measurement of attitudes. New York: McGraw-Hill.
- Simpson, R. D., Rentz, R. R., & Shrum, J. W. (1976). Influence of instrument characteristics on student responses in attitude assessment. Journal of Research in Science Teaching, 13, 275-281.
- Spielberger, C. D., Gorsuch, R. R., Lushene, R. E. (1970). Test manual for the State-Trait Anxiety Inventory. Palo Alto, CA: Consulting Psychologists Press.
- Stricker, L. J. (1969). "Test-wiseness" on personality scales. Journal of Applied Psychology Monographs, 53 (3, Pt. 2).
- Symonds, P. M. (1931). Diagnosing personality and conduct. New York: Appleton Century.
- Thacker, J.W., Fields, M. W., & Tetrick, L. (1989). The factor structure of union commitment: n application of confirmatory factor analysis. Journal of Applied Psychology, 74, 228-232.
- Throne, F. C. (1978). Methodological advances in the validation of inventory items, scales, profiles, and interpretation. Journal of Clinical Psychology, 34 (2), 283-301.

- Thurstone, L. L. (1928). Attitudes can be measured. American Journal of Sociology, 33, 529-554.
- Tittle, C. R., & Hill, R. J. (1967). Attitude measurement and prediction of behavior: An evaluation of conditions and measurement techniques. Sociometry, 30, 199-213.
- Towne, D. C. (1967). Influences exerted upon subject responses by the response scale structured elements of attitude scales. Dissertation, Cornell University, Ithaca, NY..
- Trott, D. J., & Jackson, D. N. (1967). An experimental analyses of acquiescence. Journal of Experimental Research in Personality, 2, 278-288.
- Tryon, R. C. (1939). Cluster analysis. Ann Arbor, MI: Edwards Brothers.
- Tryon, R. C., & Bailey, D. E. (1970). Cluster Analysis. New York: McGraw-Hill.
- Violato, C., & Marini, A. E. (1989). Effects of stem orientation and completeness of multiple-choice items on item difficulty and discrimination. Educational and Psychological Measurement, 49, 287-295.
- Wason, P. C. (1961). Responses to affirmative and negative binary statements. British Journal of Psychology, 52, 133-142.
- Wesman, A. G. (1952). Faking personality test scores in a simulated employment situation. Journal of Applied Psychology, 36, 112-113.
- Wiggins, J. S. (1973). Personality and prediction. Reading, MA: Addison-Wesley.
- Wiggins, J. S. (1966). Social desirability estimation and “faking good” well. Educational and Psychological Measurement, 26, 329-341.
- Winkler, J. D., Kanouse, D. E., & Ware, J. E., Jr. (1981, August). Controlling for acquiescence response set in scale development. Paper presented at the 90th annual meeting of the American Psychological Association, Los Angeles, CA.
- Zern, D. (1967). Effects of variations in question-phrasing on true-false answers by grade-school children. Psychological Reports, 20, 527-533.

MICHIGAN STATE UNIV. LIBRARIES



31293020586321