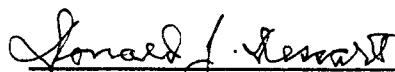


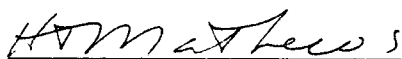
To the Graduate Council:

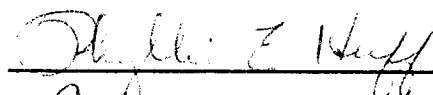
I am submitting herewith a dissertation written by Michael Scott Winstead entitled "The Development of an Instrument to Measure Two Constructs of Mathematics Teachers' Attitudes Toward Teaching Statistical Concepts." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Education.

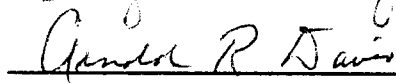


Donald J. Dessart, Major Professor

We have read this dissertation
and recommend its acceptance:







Accepted for the Council:



Associate Vice Chancellor
and Dean of The Graduate School

**The Development of an Instrument to Measure
Two Constructs of
Mathematics Teachers' Attitudes Toward
Teaching Statistical Concepts**

A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Michael Scott Winstead

August 1996

Copyright © Michael Scott Winstead, 1996
All rights reserved

DEDICATION

This dissertation is dedicated to my parents

Gerald and Jean Winstead

and my wife

Sonya Winstead

who have given me plenty of love and support.

ACKNOWLEDGMENTS

I am grateful to Dr. Donald J. Dessart, my major professor and a fine gentleman, for his many encouraging words and patience throughout my ordeal. I would also like to thank the other committee members, Dr. Phyllis Huff, Dr. Arnold Davis, and Dr. Tom Mathews, for their assistance and support.

I would like to express my thanks and love to my wife, Sonya, for remaining confident in me when my own confidence was wavering. I would like to thank my parents for showing me the importance of an education. I would like to thank my children: Kerri, Matthew, and Reagan, for showing me each day the truly important things in life. Finally, I would like to thank my Heavenly Father, without whom all else would be meaningless.

ABSTRACT

The Teachers' Attitude Toward Statistics and Probability (TATSP) scale was developed to measure secondary mathematics teachers' attitudes toward teaching statistical topics in the regular high school curriculum. The scale was based on Fishbein's theory of reasoned. The relationship between a teacher's beliefs about teaching statistics, attitudes about teaching statistics, and behaviors involving teaching statistics were consistent with Fishbein's model. The 24-item TATSP scale consists of two subscales to measure teachers' perceived importance of and liking of teaching statistical topics.

A 43-item pilot TATSP scale was constructed to solicit evaluative beliefs about either the importance or liking of teaching statistical topics, and tested with 32 high school mathematics. The scale was reduced to 24-items and tested with 58 teachers from 11 public high schools.

The TATSP scale was reliable and valid. The internal consistency estimate was 0.92 for the scale and 0.85 and 0.94 for the importance and liking subscales, respectively. Test-retest correlations were all satisfactory.

Construct and criterion-related validity were established. Correlation of total scale score with self-reported coverage of statistical topics was 0.55. About 72 percent of the teachers were correctly classified as members of the low or high coverage group using scale scores.

It was concluded that teachers who like statistics cover more statistical

topics in their classrooms. Also, high school mathematics teachers who are knowledgeable of statistics cover more topics in their mathematics classes. Teachers who graduated from college recently had completed more statistics courses than their more experienced counterparts.

TABLE OF CONTENTS

CHAPTER	Page
1. INTRODUCTION	1
Background for the Study	1
Statement of the Problem	3
Purpose of the Study	5
Importance of the Study	6
Limitations	7
Definitions	7
2. REVIEW OF LITERATURE	10
Theories of Attitude	10
Definitions of Attitude	10
The Theory of Reasoned Action	11
Teachers' Attitudes and Instructional Practices	13
Measuring Attitudes	14
Methods of Scale Construction	15
Guttman's Scalogram Analysis	15
Osgood's Semantic Differential Technique	16
Thurstone's Method of Equal-Appearing Intervals	16
Likert's Method of Summated Ratings	17
Steps in constructing Likert scales	17
Comparison Between Self-Report Methods	19
The Likert Scale and the Theory of Reasoned Action	20
Reliability and Validity of Scales	20
Types of Reliability and Their Measurement	21
Test-retest	21
Equivalent-forms	22
Split-half	23
Coefficient alpha	24
Summary	25
Validity of Attitude Measures	25
Content validity	25
Construct validity	26
Criterion-related validity	28
Relationship Between Reliability and Validity	29
Existing Statistics and Probability Attitude Scales	29
Existing Math Attitude Scales	32
3. THEORY AND HYPOTHESES	38

Theoretical Basis of the TATSP Scale	38
Hypotheses	40
Research Question	41
4. METHODS AND PROCEDURES	43
Preparation and Administration of the Pilot Scale	43
Preparation of Pilot Scale	43
Preparation of Teacher Profile Questionnaire	44
Administering the Pilot Scale	44
Selecting the Sample	44
Data Collection for the Pilot Scale	45
Preparation of the Final Scale	46
Analysis of Data from Pilot Scale	46
About the Sample	46
Scoring the Pilot Scale	46
Distribution of responses	47
Scoring of Personal Profile Questionnaire	55
Correlation of Scale Scores and Questionnaire Variables	55
Estimates of Internal Consistency	57
Selection of Items for the Final Scale	57
Administering the Final Scale	62
Selecting the Sample	62
Data Collection for the Final Scale	62
5. RESULTS	63
Analysis of Data from Final Scale	63
About the Sample	63
Distribution of Responses	63
Scoring of Personal Profile Questionnaire	67
Analysis of Data for Research Hypotheses	74
Correlation of Variables	75
Correlation of scale scores and questionnaire variables	75
Item Analysis	76
Predicting Usage and Discriminating Between Low and High Users	79
Reliability	80
Estimates of Internal Consistency	80
Test-Retest Reliability	80
6. CONCLUSIONS AND RECOMMENDATIONS	81
Description of the TATSP Scale	81
Reliability	81

Validity	82
Construct Validity	82
Criterion-Related Validity	83
Summary of Reliability and Validity	83
Factors Associated with Coverage of Statistics	84
Conclusions	85
Recommendations	85
Suggestions for Further Study	85
Studies Using the TATSP Scale	86
REFERENCES	88
APPENDICES	94
A - Cover Letter for Administration of Pilot TATSP Scale	95
B - Pilot TATSP Scale	97
C - Cover Letter for Administration of Final TATSP Scale	101
D - Final TATSP Scale	103
VITA	106

LIST OF TABLES

Table	Page
1 Distribution of responses of items on 43-item pilot scale	48
2 Summary statistics for 43-Item pilot scale after reversing scores of negative items	50
3 Total score, subscale scores, and coverage ranked by total score for pilot scale	52
4 Comparison of coverage, subscale, total scores for pilot scale by teaching experience and statistics education	54
5 Numerical scoring and distribution of responses for items on the Personal Profile Questionnaire used with the pilot scale	56
6 Correlations of scale scores and questionnaire variables for the 43-item pilot scale	58
7 Estimates of internal consistency as measured by Cronbach's coefficient alpha for 43-item pilot scale and subscales	58
8 Correlation of item and subscale scores and discriminating quality of items on 43-item pilot scale	60
9 Distribution of responses of items on 24-item final TATSP scale	65
10 Summary statistics for 24-Item Final TATSP Scale after reversing scores of negative items	66
11 Total score, subscale scores, and coverage ranked by total score for the 24-item final TATSP Scale	68
12 Comparison of total and subscale scores by coverage group	71
13 Comparison of coverage, subscale, total scores for 24-item final TATSP scale by teaching experience and statistics education . .	72
14 Numerical scoring and distribution of responses for items on the Personal Profile Questionnaire used with the final scale	73

15	Correlations of scale scores and questionnaire variables for the 24-item final TATSP scale	77
16	Item-subscale and item-coverage correlations and discriminating quality of items on 24-item final TATSP scale	78
17	Test-Retest correlation coefficients	80

CHAPTER 1

INTRODUCTION

Background for the Study

Most everyone is bombarded daily with statistical arguments. From a public opinion poll to the latest government figures on unemployment, some understanding of the results to figure out their validity and usefulness is essential. Thus, statistics and probability topics, or stochastics, are appropriate for the school mathematics curriculum (Auzmendi, 1991; Grey, 1983). Not only do they help cipher good statistical arguments from bad ones, they help consumers distinguish sound uses of statistical procedures from unsound or deceptive uses. With their practical uses in the real-world, statistics and probability topics provide meaningful applications of mathematics (Shulte, 1981). Furthermore, they are interesting, exciting, and motivating topics for most students (Mendoza & Swift, 1981; Shaughnessy & Bergman, 1993).

The National Council of Teachers' of Mathematics (NCTM) recognizes the importance of these topics. Its 1981 Yearbook dealt with teaching statistics and probability (Shulte, 1981). Stochastic topics also received much attention in the Curriculum and Evaluation Standards for Teaching Mathematics (NCTM, 1989). The authors felt that these topics were appropriate for children in all grade levels. For example, children in grades K-4 should: (1) collect, organize, and

describe data; (2) construct, read, and interpret displays of data; (3) formulate and solve problems that involve collecting and analyzing data; and (4) explore ideas of chance. Furthermore, the curriculum must emphasize that statistics is more than reading and interpreting graphs. It is describing and interpreting the world with numbers, and it is a tool for solving problems (NCTM, 1989).

Publishers and authors of mathematics textbooks have heard the call for increased emphasis on stochastic topics. Nearly all elementary mathematics textbooks now have sections or even chapters devoted to such ideas. For instance, the Heath Mathematics Connections third grade textbook has a 29 page chapter dealing with probability and statistics. The authors introduce descriptive statistics techniques, pictographs, bar graphs, and line graphs, along with basic probability concepts (Manfre, Moser, Lobato, and Morrow, 1992). A sixth grade textbook devotes 38 pages to stochastics, including more advanced probability concepts such as independent events. Furthermore, the authors discuss the idea of an average and its various measures: mean, median, and mode (Burton, Hopkins, Johnson, Kaplan, Kennedy, and Schultz, 1992). Many high school textbook authors also devote several pages to stochastics. The Heath Algebra I book contains at least one section in each chapter that focuses on probability and statistics topics such as constructing and interpreting graphs, scatter plots, and correlation and regression (Larson, Kanold, and Stiff, 1993).

There is perhaps no other branch of the mathematical sciences as important for all students, college bound or not, as probability and statistics

(Mendoza & Swift, 1981; Shaughnessy & Bergman, 1993). All students eventually become consumers and citizens at some level, and will enter a society in which the use of data and graphs to communicate information and influence decisions is ever increasing. Consequently, stochastics must enter the mainstream of school mathematics. As stated in the preface to the Proceedings of the First International Conference on Teaching Statistics, "We can't afford not to" teach statistics (Grey, 1983, p. 7).

Statement of the Problem

Although the call for increased instruction of probability and statistics has been clear, classroom teachers have been slow to respond. The responsibility of teaching the necessary stochastics to children falls on the teachers.

Shaughnessy (1992) pointed out three fundamental barriers for improvement of stochastics teaching: (1) getting stochastics into the mainstream of the mathematical science curricula at all, (2) enhancing teachers' background and conceptions of probability and statistics, and (3) confronting students' and teachers' beliefs about probability and statistics.

Teachers determine what is taught in schools. They decide (a) how much time is allocated to a topic; (b) what topics are taught; (c) what topics are taught to which students; (d) when and in what order each topic is taught; and (e) to what standards of achievement a topic is taught (Schille et al., 1982).

Usiskin (1981) identified four broad categories for inclusion of a topic or

concept in the curriculum: (1) the teacher's perceived importance to students' lives outside school, (2) the teacher's perceived importance to students' later lives in school, (3) general transfer, and (4) motivation for teacher and student. The first two categories deal with the teacher's beliefs about the importance of a topic to students. It is by far the most common rationale offered by teachers for inclusion of a topic in their curriculum. Another compelling reason falls under category four: the teacher enjoys teaching the concept or topic.

Thus, differences among teachers in the content of their mathematics instruction are partially a function of differences in beliefs and attitudes about mathematics. Although there are external influences on their curriculum, teachers will primarily look to their own beliefs to decide the content of their instruction. There is variety of curricula in schools mostly because teachers differ in their beliefs about the importance of mathematics and the most important topics within mathematics and in their interest and enjoyment in teaching mathematics (Porter, Floden, Freeman, Schmidt, & Schwille, 1988).

Teachers' attitudes are not only influential in curriculum decisions. They also play a big role in student achievement. Students can perceive and be influenced by a teacher's attitudes such as excitement, curiosity, or anxiety toward mathematics. In fact, according to Aiken (1970), "Of all the factors affecting student attitudes toward mathematics, teacher attitudes are viewed as being of particular importance" (p. 592). Furthermore, many educators and psychologists believe mathematical achievement is affected most by student

attitudes toward their ability to do mathematics well as well as the value they place on mathematics in general (Brophy, 1986; Confrey, 1986; Schunk, 1985; Buchanan, 1987; Frary & Ling, 1983). It is logical that a highly positive teacher attitude toward stochastics would stimulate student achievement in such topics.

Thus, teachers' attitudes play an important role in the classroom. They influence curriculum decisions and affect student attitudes and achievement. Consequently, a scale which measures teachers' attitudes toward teaching statistical and probabilistic topics is needed.

Purpose of the Study

The purpose of this study was to develop a scale for measuring two constructs of teacher attitude toward teaching probability and statistics topics: their perception of the importance of teaching such topics and their liking for teaching the topics. A review of research indicates that no such instrument exists. Such an instrument could be effectively and efficiently used by a variety of educators to promote academic achievement in stochastics. Teacher training personnel could use it to modify their programs so that stochastics receive more emphasis. Also, it would supply administrators an opportunity to gauge their teachers' attitudes and possibly develop inservice programs in an attempt to improve attitudes.

Importance of the Study

Affect is an important factor in teaching and learning mathematics. The NCTM *Curriculum and Evaluation Standards* put considerable emphasis on affective issues (1989). Two of the major goals of the *Standards* deal with helping students understand the value of mathematics and with developing student confidence. In this light, student attitude toward stochastics is very important.

Aiken (1974) concluded that teachers are viewed as the prime determiners of both student attitude and performance in mathematics. In assessing student attitude toward mathematics and reasons for dislike of the subject, Quilter and Harper (1988) reported that students' identified the mathematics teacher as the greatest source of dissatisfaction with the subject.

Teachers' attitudes also have a profound influence on behavior. In drafting a causal chain, Fishbein and Ajzen (1975) view beliefs as the receiver of available information needed for the formation of attitudes which in turn influence intentions which are the basis for decisions that lead to action. Thus, teachers' attitudes influence what topics are taught, when they are taught, and which instructional practices are used.

In this age of information and technology, an ever-increasing need exists to understand how information is processed and translated into usable knowledge. Because of society's expanding use of data for prediction and decision making, it is important that students develop an understanding of the

concepts and processes used in analyzing data. A knowledge of probability and statistics is necessary if students are to become intelligent consumers who can make critical and informed decisions.

Since teacher attitude plays such an important part in the students' learning any topic, an instrument is needed to measure teachers' attitudes toward teaching stochastics. There apparently is no measurement scale currently available. Consequently, the need for such a scale is essential.

Limitations

1. This study was limited to secondary school mathematics teachers in East Tennessee school systems.
2. The selection of teachers asked to participate in this study was not random. Participating teachers were chosen via the cluster sampling technique from school systems willing to participate in the study. In other words, schools were randomly selected and then all the mathematics teachers in those schools were asked to participate.
3. As with any study that relies on a questionnaire for data collection, some teachers refused to participate.

Definitions

Attitude: An attitude is considered to be the amount of affect for or against some object (Fishbein & Ajzen, 1975, p. 11). Here, affect refers to a

person's feelings toward and evaluation of the object. This definition restricts the attitude concept to an unidimensional component. This suggests that in order to measure an individual's attitude toward a given object a procedure should be used which will locate the subject on a bipolar affective or evaluative dimension (Turner, 1987, p.40).

Construct validity: An index of the degree to which an individual's performance on a measuring instrument can be ascribed to the construct under consideration (Lemon, 1973). The researcher validates his scale by investigating whether it confirms or denies the hypothesis predicted from a theory which is based on the construct (Bohrnstedt, 1970).

Content validity: The degree to which the scale being used represents the concept about which generalizations are to be made (Bohrnstedt, 1975, p. 91).

Importance of teaching stochastics: An individual is said to perceive the importance of teaching stochastics when he indicates that he feels that stochastics learned in school is useful later in life. He believes that statistical concepts are important for getting a job or are useful on the job. He feels that every person should know a minimum amount of stochastics to function efficiently in the everyday world. An individual is also said to perceive teaching stochastics as important if he indicates such concepts as useful in later life in school. In other words, the concepts appear in other mathematics courses or on college entrance tests.

Liking of teaching stochastics: An individual is said to like teaching stochastics when he indicates that he views stochastic topics as some of his favorite topics to teach. They are seen as enjoyable, interesting, and fun to teach. Also, the concepts are viewed as fun or encouraging to students to learn more.

Reliability: Refers to the degree to which a measure is free of variable error. Thus reliability concerns the extent to which measurements are consistent.

Stochastics: Commonly used to jointly refer to statistics and probability.

TATSP (Teachers' Attitudes Toward Statistics and Probability) Scale: The scale developed in this study for the purpose of measuring secondary mathematics teachers' attitudes toward teaching statistical and probabilistic topics.

Validity of an attitude scale: The degree to which the scale measures what it is supposed to measure (Shaw and Wright, 1967, p. 18).

CHAPTER 2

REVIEW OF LITERATURE

Theories of Attitude

Definitions of Attitude

The definition of attitude has varied in meaning from one investigator to the next. Presently, there is still no one universally accepted definition of the attitude concept. For some it is a unidimensional construct highly correlated, but separate, from beliefs and behavioral intentions. Others view the attitude concept as multidimensional with beliefs and intentions included in the construct.

One of the major early advances in attitude research was made by Thurstone. In Thurstone's model, attitude denoted the total accumulation of a person's feelings, biases, ideas, fears, and convictions about any social object. He defined attitude as "the affect for or against a psychological object" (Thurstone, 1931).

Doob (1947) investigated the relationship between attitude and behavior. In essence, he attempted to explain the failure of attitudes to predict behavior. He defined the attitude concept in behavioral terms as a learned implicit response which is anticipatory and mediating in references to patterns of overt responses and is socially significant. In formulating his definition in this manner,

Doob attempted to show that attitude is a quasi-scientific term which, to be understood, must be subsumed into more basic theory, namely behavior theory.

Another approach to explain the failure of attitudes to predict behavior involved viewing attitude as a multicomponent construct. Krech and Crutchfield (1948) made explicit this characterization of attitudes writing, "Beliefs can be seen as the cognitive embodiment of attitudes. All attitudes incorporate relevant beliefs" (p. 152). Further, Krech and Crutchfield argue that attitude is characterized by a "driving force". Thus attitudes have components which are cognitive and active.

Katz and Scotland (1959) defined attitude as a predisposition to evaluate an object in a certain way. This evaluation is composed of three components: affective, cognitive, and behavioral. They viewed the affective component as the central aspect of the attitude. They characterized the cognitive component by the number of cognitive elements or beliefs. Katz and Scotland defined the behavioral components as an action tendency toward the object of the attitude.

The Theory of Reasoned Action

The theory of reasoned action (Fishbein & Ajzen, 1975) is a conceptual framework which both unifies and delineates the various constructs involved in attitude research. The theory involves four major concepts -- beliefs, attitudes, intentions, and behaviors -- and the relationship between them. In brief, the theory holds that a person's behavior is immediately determined by his intention

to perform the action. The person's intention is a function of his attitude toward the behavior and his subjective social norm -- his beliefs about what significant referents think -- relative to the behavior. The attitudes and subjective norms of the individual are each functions of his beliefs. As a foundation for this theory, the assumption is made that human beings are normally rational and use the information available to them to make judgments, form evaluations, and arrive at decisions.

Like most other investigators, Ajzen and Fishbein agree that attitude can be described as a learned predisposition to respond consistently in a favorable or unfavorable manner with respect to a given object (Fishbein, 1967; Fishbein & Ajzen, 1975). However, a distinguishing feature of their theory of reasoned action is the separation of beliefs from attitudes, attitudes from behavioral intentions, and intentions from overt behavior. Their model relates an individual's attitudes to his behavior by means of his intentions to behave in a certain fashion.

Beliefs are the fundamental building blocks in Fishbein's model. A person learns or forms a number of beliefs about an object. In other words, he associates the object with various attributes. The totality of a person's beliefs serves as the informational base that ultimately determine his attitudes, intentions, and behaviors.

An individual's attitude toward an object X is a function of his set of beliefs about X. If those beliefs associate the object with primarily favorable attributes,

his attitude will tend to be positive. Conversely, negative attributes will result in a negative attitude. Thus, a person's attitude toward some object is determined by his beliefs that the object has certain attributes and by his evaluations of those attributes. An algebraic expression of the relationship between the beliefs and attitude is the following (Fishbein & Ajzen, 1975).

$$A_x = \sum_{i=1}^n e_i b_i \quad (1)$$

where

A_x = attitude toward object X,

b_i = belief "I" about X,

e_i = evaluative aspect of b_i ,

n = number of salient beliefs.

Teachers' Attitudes and Instructional Practices

Numerous studies have been conducted to investigate the relationship between the attitudes of teachers and their instructional practices. More precisely, the impact attitudes have on the content of mathematics instruction. In considering the content it is useful to distinguish between what is taught and how much is taught. "What" pertains to the substance of topics or concepts; "how much" concerns the number of different topics treated (Dreeben, 1985).

In his study of fourth-grade teachers, Barr (1988) found the textbook to be a strong influence on the nature and sequence of topics covered. However, he

reported that any deviations from the textbook presentation could be traceable to teachers' beliefs about the subject matter, including the importance of certain topics over others. Topics not highly valued by the teachers such as difficult division and geometry, were either not introduced or were covered in a superficial way.

Usiskin (1981) found the most compelling reason for including a concept or topic was the teacher's perceived importance of the topic to students. He also found teachers inclined to teach topics that they found interesting or enjoyable to teach.

Measuring Attitudes

Since attitude is an 'inferred' construct there can be no such thing as a direct measure of attitude. Information about attitudes can be gathered either by direct observation of behavior or by the self-administered questionnaire (Lemon, 1973, pp. 55-82). In the final analysis, all measurement involves observation of one or more responses made by a subject, whether they are verbal or overt behavioral responses. Much research in the attitude area has relied on single-response measures to infer beliefs, attitudes, and intentions (Fishbein and Ajzen, 1975, pp.53-89). Anderson (1981, pp. 55-74) concluded that self-report methods were more appropriate than direct observation on both practical and theoretical grounds.

Methods of Scale Construction

There are several methods for attitude scale construction. However, there are only four well-known methods: Likert's methods of summated ratings; Thurstone's methods of equal-appearing intervals; Guttman's Scalogram analysis; and Osgood's Semantic differential technique.

Guttman's Scalogram Analysis

A major objective of the Guttman scaling procedure is to demonstrate that responses fall into a cumulative pattern. A perfect cumulative scale has the following condition: A person solves all items up to a certain difficulty level and no items beyond that level.

In this method, the investigator begins by obtaining a number of belief items that are related to the attitude object under consideration. First, responses are obtained from a sample of subjects, and the items are ordered in terms of their frequencies of endorsement. Once the items have been ordered, the obtained response patterns of each subject can be examined. If the scale is cumulative, a subject who endorses an item of given favorableness should also respond to all less-extreme items (Crano and Brewer, 1973). Under an ideal condition, the knowledge of a subject's total score should enable the investigator to reproduce accurately the individual's pattern of responses, a quality known as reproducibility (Oppenheim, 1966).

Osgood's Semantic Differential Technique

Osgood's technique (1971) requires a listing of bipolar adjective (e.g. fair-unfair) to which subjects respond to one of seven points on a continuum between each set. When the data of semantic differential are subjected to a factor analysis, three factors of meaning emerge: evaluation, potency, and activity. Osgood (1957) argued that a person's attitude toward an object is equivalent to the object's evaluative meaning for the person. Thus, in defining the respondent's attitude, only evaluative scores are summated (Crano and Brewer, 1973).

Thurstone's Method of Equal-Appearing Intervals

Thurstone's method was the first method of attitude scaling to be developed. The first step in Thurstone scaling involves the collection of a large pool of belief items related to some attitude object. Thurstone (1931) assumed that responses to such items are expressions of the person's attitude. More specifically, he made the assumption that different items may express different degrees of favorableness or unfavorableness toward the attitude object. A major purpose of Thurstone scaling is to specify the location of each item on the evaluative dimension by assigning a scale value to the item.

The pool of items are presented to a sample of judges representative of the population of subjects whose attitudes are to be assessed. The judges then arrange each statement over an 11-point scale from extremely favorable to

extremely unfavorable. The mean rating by the judges is considered the point for that item along the 11-point continuum. About 20 items are selected for the final scale such that (1) each item has a small standard deviation and (2) the mean ratings spread evenly at points along the continuum. In administration of the scale subjects are instructed to mark only those statements with which they agree. The score for the subject is the average scale value of the items endorsed (Nunnally, 1978).

Likert's Method of Summated Ratings

Likert's method is based on a linear model. Thus, items on a Likert scale will not be cumulative. The model assumes that individual items are monotonically related to underlying attitudes and that a summation of item scores is approximately linear related to the attitude. One obtains a total score by adding scores of individual items.

Steps in constructing Likert scales.

Eight steps are needed to construct Likert scales, an adaptation of those by Anderson (1981, pp. 248-252).

1. The first step in constructing a Likert scale is to collect a large pool of statements of either beliefs or intentions.
2. Judges classify each item as expressing positive (favorable), negative or neutral attitude toward the object in question. Judges should be from the

audience for whom the scale is intended.

3. Items classified as neutral and those that are not unanimously classified as positive or negative are eliminated.

4. The number of response choices for each item is determined. The original Likert scale had five response categories: strongly agree, agree, not sure, disagree, strongly disagree. Modifications allow for either more or fewer choices.

5. The instrument is prepared with instructions to respondents that they should indicate how they feel about each item by marking strongly agree, etc.

6. The preliminary scale is administered to a sample of the audience for whom the instrument is intended. Five-response choice items are scored from 1 to 5. Strong agreements with favorable items are given a score of 5, and strong disagreements with these items are given a score of 1. Scoring is reversed for unfavorable items, such that disagreement with an unfavorable item results in a high score. The person's attitude score is obtained by summing across all his item scores. The higher the score, the more favorable the attitude.

7. The next stage is one of item analysis. The task is to select those items which are best able to differentiate between the respondents. Two methods are commonly used to accomplish this task. One is item-total correlations. Correlations between responses for each item and the total scale score are computed. Those items which demonstrate the highest correlations are chosen for the final scale. Another method is to select the respondents with

the highest total scores and an equal number with the lowest. Mean item scores are computed for each group and tested for statistical significance. An item is chosen for the final scale if it discriminates between the extreme groups and thus has a statistically significant difference between mean scores.

Correlational techniques are more commonly used but the two methods generate similar results (Murphy and Likert, 1937).

Comparison Between Self-Report Methods

The Likert scale has a number of advantages over all other methods which have helped to make it popular. Crano and Brewer (1973) compared Likert with Thurstone and Guttman techniques in terms of reliability and ease of construction. They wrote the following:

The Likert model of scale construction represents currently the most popular approach to the generation of reliable attitude measurement devices. When compared with either the Thurstone or the Guttman method, Likert's scale construction technique typifies a process that is not only more efficient in terms of resource expenditure but also more effective in developing scales of higher reliability. (P. 239)

Nunnally (1978) reported that summative scales have a number of attractive advantages over all other methods: (1) they follow from an appealing model, (2) they can be adapted to the measurement of many different kinds of attitudes, (3) they provide meaningful results in many studies up to date, and (4) they are rather easy to construct and highly reliable.

The Likert Scale and the Theory of Reasoned Action

All of the attitude-scaling procedures discussed arrive at a single attitude score based on responses to statements of beliefs or intentions. All scaling procedures locate individuals along an evaluative dimension (by assigning attitude scores). Generally speaking, in all procedures subjects agree or disagree with items that indicate some degree of favorableness or unfavorableness toward the attitude object. A score is obtained for each item by multiplying degree of agreement and item favorableness. This item score is equal to the product of belief about the object (b) and the evaluation of the object's attributes (e) as in Fishbein's model (Fishbein & Ajzen, 1975).

The differences between scaling procedures concerns the relative weights placed on beliefs and evaluations of certain attributes. For the Likert scale, each item is judged to indicate either a favorable (1) or unfavorable attitude (-1). Responses are given on a five-point scale ranging from strong agreement (2) to strong disagreement (-2). Thus belief strength varies from -2 to 2, and evaluations of associated attributes are either +1 or -1. Greater weight is placed on b than on e in computing attitude scores.

Reliability and Validity of Scales

The concepts of reliability and validity concern the degree to which the measuring instrument is free of measurement error. An observed score can be viewed as composed of a "true" score and some measurement error. The error

component can be further divided into two parts: variable or random error and constant error. This model is presented in equation 2,

$$x_i = t_i + e_v + e_c \quad (2)$$

where x_i is the observed score, t_i is the "true" score, e_v is the variable error, and e_c is the constant error.

Types of Reliability and Their Measurement

Reliability refers to the degree to which a measure is free of variable error. Thus reliability concerns the extent to which measurements are consistent. Bohrnstedt (1970) wrote:

What is meant by reliability? Perhaps the best synonym is consistency. If no true change occurs in a given attitude an individual holds, does the attitude scale consistency yield the same ordering for him relative to the others? If not, the scale is unreliable (p. 83).

There are four numerical methods of estimating the reliability of an attitude scale: (1) the test-retest, (2) the equivalent-form, (3) the split-half, and (4) coefficient alpha.

Test-retest.

This method of estimating reliability corresponds most closely to the conceptual notion of reliability, and the procedure is simple. The attitude scale is administered to the same group of persons at two different times, and the

correlation between the two sets of scores is computed. This coefficient, usually the Pearson r , is the reliability estimate.

Shaw and Wright (1967) reported that "the test-retest has the advantage of holding constant the items used, thus eliminating unreliability due to differences between items, which occurs in the equivalent-forms method. It also has the advantage of requiring only a single scale; no additional forms are needed" (p. 16).

According to Lemon (1973), the test-retest method of assessing reliability also has a number of disadvantages. The primary one is the influence that the first administration of the measure is likely to have on the second. It may influence the respondent's performance on the second administration because he remembers his previous response. On the other hand it may force the respondent to consider issues which he had not considered before, and thus leading to attitude change.

Equivalent-forms.

This method of estimating reliability requires two forms that may be considered equivalent. The two forms are then administered to a group of subjects, and the two sets of scores are correlated to obtain an estimate of reliability. The major advantage of this method is that the effects of time interval and of responding to one scale upon the response to the other are minimized (Shaw & Wright, 1967).

A disadvantage of this method is that at least two forms are necessary and that the correlation coefficient reflects not only consistency of measure, but also the degree to which the two forms actually do measure the same attitude. Therefore, the acceptability of this estimate of reliability depends upon the degree of equivalents of the two forms (Shaw & Wright, 1967). According to Bohrnstedt (1970), "it is practically impossible to find two completely parallel sets of items" (p. 87).

Split-half.

This method estimates reliability by treating each of two parts of the attitude scale as a separate scale. The scales may be selected by treating odd-numbered items as one scale and even-numbered items as another, or by randomly selecting items to constitute the subscales. Regardless of the method of choosing the subscales, the reliability estimate is the correlation between the scores on the separate scales. Since the resulting scales are shorter than the original scale, and because reliability tends to increase with scale length, the correlation is corrected for length by the Spearman-Brown prophecy formula $R = 2r / (1 + r)$ where R is the corrected reliability for the whole test and r is the reliability for a half test (Garrett, 1966, p. 339).

The split-half method has a number of disadvantages. The primary one is its lack of uniqueness. Cronbach (1951) wrote:

Instead of giving a single coefficient for the test, the procedure gives

different coefficients depending on which items are grouped when the test is split into two parts. If one split may give a higher coefficient than another, one can have little faith in whatever result is obtained from a single split. Such a coefficient is a property of a pair of tests, not a single test (p. 298).

Again, because of the arbitrary way in which the halves are chosen, the case for widespread use of the split-half in attitude measurement is not strong (Lemon, 1973).

Coefficient alpha.

On the meaning of coefficient α , Cronbach (1951) wrote:

A general formula (α) of which a special case is the Kuder-Richardson coefficient of equivalence is shown to be the mean of all split-half coefficients resulting from different splittings of a test. α is therefore an estimate of the correlation between two random samples of items from a universe of items like those in the test. α is found to be an appropriate index of equivalence (p. 297).

Nunnally (1978) defined coefficient α as the expected correlation of one test with an alternative form containing the same number of items. According to him coefficient α provides a good estimate of reliability in most situations, since the major source of measurement error is because of the sampling of content.

A formula for coefficient α is given by Cronbach (1951, pp. 299-305):

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum v_i}{v_t} \right) \quad (3)$$

Here v_t is the variance of test scores, and v_i is the variance of item

scores. From the above formula it is clear that α increases as the number of items increases.

Summary.

There is abundant evidence that standard attitude scales are highly reliable, yielding comparable results when administered on different occasions (Fishbein & Ajzen, 1975). It is not unusual to find test-retest and alpha coefficients at levels of .80 and .90 for Thurstone and Likert-type scales (Aiken, 1980).

Validity of Attitude Measures

The validity of a test, or any measuring instrument, depends on the extent to which it measures what it was to designed to measure. Evidence for the validity of attitude measures is of three kinds: content validity, construct validity, and criterion-related validity (Aiken, 1980).

Content validity.

Content validity is defined by Bohrnstedt (1970) as the degree to which the scale being used represents the concept about which generalizations are to be made (p. 91). Evidence for content validity is obtained from careful inspection of the scale items by "experts" and their judgments of its validity.

Construct validity.

A construct, as its name implies, is something that is derived from a number of different observations. It is an inference made by the investigator to explain the relationships he discovers between different observations. Attitude is a construct of this kind. A construct is therefore an abstraction which cannot be defined by any one measure, but which enters in with other factors to determine performance in a variety of measures. The construct validity of a particular measure is an index of the degree to which an individual's performance on that measure can be ascribed to this construct (Lemon, 1973). The researcher validates his scale by investigating whether it confirms or denies the hypothesis predicted from a theory which is based on the construct (Bohrnstedt, 1970).

There are several methods of estimating construct validity. The known-group approach and the convergent and discriminant validation technique are the most recommended.

The known-group approach is the most familiar approach (Crano and Brewer, 1973). In this validation technique various groups of persons known to differ on an attitude tapped by the scale under investigation are administered the scale. The reasoning behind this technique is simple. If the scale is really a measure of the particular attitude in question, groups known to be different should generate clearly discriminable mean scores.

For example, if one wants to verify that a scale designed to measure

teachers' attitudes toward teaching statistics has construct validity by using the known-group approach, the scale will then be administered to two different groups of teachers: (a) a group of teachers who show high interest in statistics by choosing to teach statistical topics; and (b) a group of teachers who reject teaching statistical topics. For the scale to have construct validity, the mean score of the first group should be much higher than the mean score of the second group.

A series of operations, termed convergent and discriminant validation techniques by Campbell and Fiske (1959), are frequently employed in the assessment of scale validation. In convergent validation, an investigator is asked to generate a series of hypotheses regarding probable interrelationships that might exist between his construct and others. If this scale is a valid indication of the construct under study, and the relationships that he has hypothesized do in fact exist, the scales assessing these various conceptualizations should prove to be highly related. In discriminant validation, he must devise a series of variables with which his construct would not be expected to relate. If the scale is discriminative, and the investigator's hypotheses are correct, no interrelationship will be expected between the values of such scales (Crano and Brewer, 1973).

As an example, if an attitude scale that measures teachers' attitudes toward teaching statistics has construct validity, a correlation between this scale and a scale that measures teachers' attitudes toward reading should be very

low; it has discriminant validation. Also a correlation between a scale that measures teachers' attitudes toward mathematics and the same scale should be positive, and even moderately high; it has convergent validation.

Criterion-related validity.

Information about the criterion-related validity of an attitude instrument is obtained by correlating attitude scores with scores on the same or another variable obtained with another measure at the same time (concurrent validity) or at a later date (predictive validity). Normally, predictive validity is used when the purpose of the scale is to predict some form of behavior (Crano and Brewer, 1973). The behavior itself becomes the criterion, and the attitude score is the predictor. Generally, predictive validity is established through the use of a coefficient of correlation. This coefficient is an index of the relationship between a test performance and the criterion behavior. For example, one might wish to predict school attendance from attitude toward the school. Also the number of mathematics courses could be predicted from attitude toward mathematics.

Concurrent validity differs from predictive validity, chiefly with regard to the time at which the criterion measurement is acquired. If the criterion measure and the attitude scale are administered at approximately the same time, the procedure is called concurrent validation; if the attitude score is obtained first and the criterion score at some future date, it is called predictive validity.

Relationship Between Reliability and Validity

Equation 2 above indicates that a measure cannot be valid unless it is reliable. In other words, if variable error is present, and hence the scale is unreliable, then the observed score cannot be equivalent to the “true” score. However, reliability of a scale is not a sufficient condition for validity of the scale. It is easily conceivable that a stable, internally consistent instrument could be developed that had no relationship whatsoever to the attribute that motivated the research in the first place. Consistency of responses, from item to item or from time to time, is not a guarantee of scale validity.

The reliability and validity of an instrument increase as the number of items on the scale increases, thus rendering the test a better measure of its criterion (Garrett, 1966). Reliability increases in accordance with the Spearman-Brown prophecy formula:

$$r_{nn} = \frac{nr_{11}}{1 + (n - 1) r_{11}} \quad (4)$$

where r_{nn} = the correlation (reliability) of a test increased in length by a factor of n items, and r_{11} = the reliability coefficient of the original test.

Garrett (1966) provides details of the increase in the validity coefficient (p. 356).

Existing Statistics and Probability Attitude Scales

Analysis of the literature indicates many studies undertaken to find out

attitudes among a variety of samples concerning countless areas of interest. However, the review of literature has not produced any study dealing with measuring teacher attitudes toward stochastics. Therefore, the attitudinal instruments presented in this section relate to the measurement of student attitudes toward stochastics.

Roberts and Bilderback (1980) developed a 34-item 5-point Likert Statistics Attitude Survey (SAS). The scale was to measure the attitudes toward statistics of students taking a beginning statistics course. Items tap such areas as usefulness of statistics and anxieties of students who must take statistics. The total scale scores were correlated with grades earned in a basic statistics course. The researcher found the correlations to be moderate and positive.

In a follow-up study, coefficient alpha reliability for the SAS was reported to be .92. Total scale scores were also found significantly related to such cognitive variables as basic mathematical skills, statistics preknowledge, and course grades. Non-cognitive factors with which SAS was significantly correlated were sex, the degree to which students suggested that they wanted to take the course and that they were glad they had taken the course, number of previous mathematics courses completed, the status of a course being required or elective, calculator attitudes, and course and instructor evaluations. In addition, SAS scores showed a significant positive change from the beginning to the end of the course (Roberts & Saxe, 1982).

Wise (1985) constructed the Attitudes Toward Statistics (ATS) scale to

measure the attitude change in introductory statistics students. The total scale consists of 29 5-point Likert type items. Two ATS subscales are identified: Attitude Toward Course and Attitude Toward the Field. The internal-consistency reliability, coefficient alpha, was reported to be .90 and .92 for the two subscales, respectively. Also, application of the test-retest procedure yielded reliability coefficients of .91 and .82. Scree test showed presence of two common factors corresponding to the subscales that accounted for 49% of the total variance. Woehlke (1991) supported many findings of Wise.

Roberts and Reese (1987) compared the Statistics Attitude Survey and Attitude Toward Statistics scales. The researchers combined the scales and administered it to 280 introductory statistics students. They found the correlations between the two scales to be .88 and concluded that the ATS was essentially an alternate form of the previously developed SAS.

Sutarso's (1992) Students Attitudes Toward Statistics (STATS) consists of 21 items which requires students to describe themselves based on a zero to nine scale, with zero being "does not describe me" and nine being "describes me." The researcher reported a coefficient alpha of .86. R-Factor analysis showed that there were six underlying dimensions that comprise the student's attitudes toward statistics. They were student's interest and applicability; relationship and impact of the instructor; attitude toward statistical tools; self confidence; parental influence; and initiative and extra effort in learning statistics.

Auzmendi (1991) produced the Multifactorial Scale of Attitudes Toward Statistics (MSAS). The researcher identified five possible dimensions of attitudes toward statistics for assessment: Enjoyment, Anxiety, Motivation, Confidence, and Usefulness. The scale has 25 items assessing the five different dimensions using a 5-point Likert type format. The scale was administered to 101 undergraduates in Spain with coefficient alpha between .60 and .87.

Existing Math Attitude Scales

The study of attitudes toward mathematics has been an active research area of research for the past forty years. Periodic general reviews of this literature exist (Aiken, 1970b, 1976; Kulm, 1980). This review will concentrate on the prominent instruments developed to measure attitudes toward mathematics.

The study of attitudes toward mathematics was begun in a series of studies by Dutton (1951, 1954, 1956, 1962). He asked prospective elementary school teachers to list their feelings toward arithmetic which were favorable and those which were unfavorable. After collecting these responses from more than six hundred students over a five-year period, Dutton selected 83 statements which were analyzed and edited into forty-five items. These items were then sorted and assigned a scale value from 1, indicating extreme dislike, to 11, indicating extreme liking, according to the methods of Thurstone (1928). The

final scale composed of 22 statements was administered to 289 students. Test-retest reliability was reported to be .94. Dutton and Blum (1968) produced the Attitude Toward Arithmetic Scale by taking the strongest items from Dutton's 1951 scale plus new items. They administered the 27 Likert-type items to a sample of 346 students in grades six to eight. Spearman-Brown test-retest reliability was .84.

Aiken and Dreger (1961) developed the Math Attitude Scale by reducing the comments of college students describing their attitude toward mathematics to 20 Likert-type items. Aiken (1963) refined the scale and the resulting Revised Math Attitude Scale has been used extensively and both its reliability and validity are considered satisfactory (Shaw & Wright, 1967). The items in the revised scale were designed to express a feeling toward mathematics. nine of the twenty items use the words "feel" or "feeling."

Anttonen (1968) constructed the Mathematics Attitude Scale which was composed of 15 Guttman-type scales for scoring by a developed method. He conducted a six-year longitudinal study whereby the scale was administered to students in grades 5 and 6 and grades 11 and 12. Correlation between the elementary attitude score and secondary attitude score was .31. The researcher reported that the scale yielded a consistent and reliable scoring procedure at both grade levels.

Adams and Van Brock (1967) developed the A-V Scale of Attitudes Toward Arithmetic. The final form consisted of 35 Likert-type items. It was

administered to 35 preservice elementary school teachers, yielding a split-half reliability of .90. Also, the correlation between scores on positive and negative items was .70.

Scharf (1971) used the semantic differential in measuring attitudes of elementary school children toward mathematics. On a semantic differential scale, the subject is asked to indicate his association of a given concept using a series of bipolar word pairs, or antonyms, such as Good versus Bad. He used the technique because it possesses the required properties of simplicity of format, content, and scoring, opaqueness regarding the "appropriate" responses, and sensitivity to the degrees of attitudinal intensity.

Most researchers of the 1960's viewed attitude as an unidimensional construct. Thus, the attitude-measuring instruments tended to be broad in nature. Aiken (1970a) called for the development of multidimensional instruments which could distinguish between different aspects of components of attitude toward mathematics.

Following this recommendation, Sandman (1973/74) developed and validated a multidimensional Mathematics Attitude Inventory. Noting that existing scales which were said to measure specific components of attitude toward mathematics lacked both internal and external validation, Sandman proposed to produce an instrument with "high reliability, low intercorrelations, internal validity through factor analysis, and external validity through relationships with other variables" (Sandman, 1973/74, p.23).

Sandman began with eight factors of attitude toward mathematics: perception of the mathematics teacher; anxiety toward mathematics; value of mathematics in society; self-concept in mathematics; enjoyment of mathematics; motivation in mathematics; perception of the mathematics class regarding the learning process; and perception of mathematics materials. Items for each of these factors were taken from seven previously developed instruments, including Dutton's scale and the Revised Math Attitude Scale of Aiken discussed earlier. In addition, some new items were written from responses to course-evaluation questionnaires. The ten items in each scale were placed in a Likert format.

After review and pilot-testing, the last two scales were dropped and the remaining scales were reduced to eight items each. This final form of the Mathematics Attitude Inventory was administered to 2,547 students in the eighth and eleventh grades. The goal of "high reliability" was achieved. Each scale had a reliability estimated by Cronbach's alpha coefficient greater than .75. However, the scale intercorrelations were not all sufficiently low to rule out the possibility that several scales were measuring the same component.

To clarify this problem and to establish the internal validity of each scale, a common factor analysis with principle axes factoring of the items was performed followed by an orthogonal varimax rotation. Sandman concluded from the results that several items should be omitted and others should be assigned to different scales. This was done and Sandman computed estimated factor scores for each subject. The resulting factor intercorrelations were sufficiently

low that each of the six factors could be interpreted in terms of the attitude construct which it was identified.

A second multidimensional scale was developed by Aiken (1974). Recognizing that most previous attitude scales primarily measured an "enjoyment of mathematics" dimension, Aiken formed an E Scale of 12 items to measure that dimension. In addition, eleven items were written to measure the value or importance of mathematics and formed the V Scale. The scales were written using the Likert method and administered to 185 college freshmen. Based on item-scale correlations, one item from each scale was omitted from the final scale. Each scale had a high reliability as measured by coefficient alpha, .95 for the E Scale and .85 for the V Scale. However, the correlation between the E and V Scales was .64, indicating significant overlap between them.

A factor analysis of data collected on the E and V Scales by Bowling (1976/1977) indicated that the constructs being measured by the two scales were distinct. However, a third scale constructed by Bowling to measure the "nature of mathematics" was not shown to be measuring a clearly separate construct from that measured by the E Scale.

Nolen, Archambault, and Greene (1976) administered the E and V Scales to 96 teachers. A factor analysis of data collected resulted in a three-factor solution. Essentially, the V Scale was split into the General Value of Mathematics and the Personal Value of Mathematics Scales.

A third multidimensional scale to measure attitudes toward mathematics

was developed by Fennema and Sherman (1976). The Fennema-Sherman Mathematics Attitude Scale consists of a group of five instruments: (1) Mathematics Anxiety Scale, (2) Attitude Toward Success in Mathematics Scale, (3) Effectance Motivation in Mathematics Scale, (4) Usefulness of Mathematics Scale, and (5) Confidence in Learning Mathematics Scale. Item responses for the five tests are obtained on a four-point Likert scale. Split-half reliabilities for the five tests were given by the researchers with coefficients ranging from .87 to .93. Investigators found relatively low intercorrelations among test scores obtained through administration of the five instruments mentioned above. The researchers, therefore, concluded that each scale measured a different construct.

CHAPTER 3

THEORY AND HYPOTHESES

Theoretical Basis of the TATSP Scale

The theory of reasoned action model proposed by Fishbein and Ajzen (1975) forms the theoretical framework for this proposed study of teachers' attitudes toward statistics. Fishbein and Ajzen, as has been noted, make a distinction between three constructs (belief, attitude, and behavioral intention) and provide a model for the relationship among these constructs and behavior.

Attitude was previously defined as the amount of affect for or against some object. "Attitude toward teaching stochastics" is the amount of affect for or against teaching stochastics. It is the individual's general feeling of favorableness or unfavorableness toward teaching stochastics. One objective of this study is to predict the degree to which statistical topics are covered by estimating the attitude of the teacher towards teaching such topics. The degree to which a topic is taught is influenced by the perceived importance and liking of the topic by the teacher. Thus, two constructs of attitude toward teaching stochastics, "attitude toward the importance of teaching of statistics" and "attitude of the liking of teaching of statistics", were investigated.

In order to measure the importance and liking dimensions of attitude toward teaching stochastics, the above definition requires that the procedure

locate the subject on a bipolar affective or evaluative dimension. Both of the commonly used types of scales, Likert and semantic differential, are bipolar. The crucial criterion is that all items in a Likert-type scale must refer to the subject's affect for or against statistics and all word-pairs in a semantic differential must be evaluative and again refer solely to statistics. The Likert-type scale was chosen as the basis for the measurement procedure in this study.

Fishbein's model proposes that teachers' attitudes toward teaching stochastics are a function of their beliefs about teaching stochastics. More specifically, the model indicates that attitude can be measured by summing across the collection of these beliefs and the teachers' evaluations of them as in the following algebraic expression.

$$A_X = \sum_{i=1}^n e_i b_i \quad (5)$$

where

A_X = attitude toward object X,

b_i = belief "i" about X,

e_i = evaluative aspect of b_i ,

n = number of salient beliefs.

This investigation is concerned with describing teachers' attitudes of the importance of teaching stochastics and their liking of teaching such topics. Thus, the portion of Fishbein's model expressed in Equation 5 provides the

theoretical framework for the study.

Hypotheses

The primary hypothesis of this study is that Fishbein's theory of reasoned action model is valid when the object under consideration is teaching statistics. In other words, the relationship between a teacher's beliefs about teaching statistics, attitudes about teaching statistics, and behaviors involving teaching statistics, will be consistent with Fishbein's model. The following research hypotheses are developed to support this primary hypothesis and validate the scale.

1. There is a positive correlation between the measure of the two attitudinal constructs, importance of teaching stochastics and the liking of teaching stochastics. The Pearson product moment r correlation coefficient between Importance subscale and Liking subscale scores was computed and the null hypothesis ($\rho = 0$) tested for significance at a 0.05 level of significance.
2. There is a positive correlation between teachers' intentions involving teaching statistical topics and the estimate of the importance of teaching stochastics. The Pearson product moment r correlation coefficient between Importance subscale scores and self-reports of intentions was computed and the null hypothesis tested for significance at a 0.05 level of significance.
3. There is a positive correlation between the teachers' intentions involving teaching statistical topics and the estimate of the liking of teaching

stochastics. The Pearson product moment r correlation coefficient between Liking subscale scores and self-reports of intentions was computed and the null hypothesis ($\rho = 0$) tested for significance at a 0.05 level of significance.

4. There is a positive correlation between the measure of importance of teaching statistics and teachers' actual coverage of statistical topics. The Pearson product moment r correlation coefficient between Importance subscale scores and the self-reports of coverage was computed and the null hypothesis ($\rho = 0$) tested for significance at a 0.05 level of significance.
5. There is a positive correlation between the measure of liking of teaching statistics and teachers' actual coverage of statistical topics. The Pearson product moment r correlation coefficients between Liking subscale scores and the self-reports of coverage was computed and the null hypothesis ($\rho = 0$) tested for significance at a 0.05 level of significance.

Research Question

The following research questions is established: What combination of the estimates of the two attitudinal constructs provides the best estimate of secondary mathematics teachers' instructional behavior concerning teaching stochastics topics?

The following analyses was performed in order to answer the research question.

1. Data from respondents who indicated that they have no responsibility in

choosing the topics covered in their classroom were eliminated. Next, Pearson product moment r correlation coefficients were computed for total scale scores and self-reports of statistical coverage, and the two subscales scores and coverage. The correlation coefficients were tested for statistical significance at a 0.05 level of significance and compared to determine which score has the strongest relationship.

2. Multiple linear regression was performed with the self-reported coverage of statistical topics as the dependent variable and the two estimates of the attitudinal constructs as predictors. Regression coefficients were tested for statistical significance at the 0.05 level of significance. The final regression equation were used to generate predicted levels of coverage. The Pearson product moment correlation coefficient were computed between predicted level of coverage and the reported level, tested for significance, and compared to the correlation coefficients computed above to determine if this combination of the two scores is more related to level of coverage than either score alone is.

CHAPTER 4

METHODS AND PROCEDURES

This chapter presents the procedures that were used to prepare and test the TATSP Scale. The preparation of the scale consisted of three phases. First, a pilot scale consisting of several items concerning teaching statistical topics was prepared and administered to a sample of the population under study. Next, the data were analyzed and items selected so that the final scale was understandable, valid, and reliable. Finally, the final TATSP scale was administered to a sample of teachers.

Preparation and Administration of the Pilot Scale

Preparation of Pilot Scale

Relevant concepts and topics about stochastics were identified through the researcher's experience and by a review of literature including statistics textbooks. A pool of items was constructed so that the items corresponded to the relevant concepts. The items elicited evaluative responses along the dimensions of like/dislike or important/not important since these were found to be the most compelling reasons teachers include a concept or topic. The following are some of the representative statements: "Statistical topics provide

students with real-world applications of mathematics” and “Knowledge of statistical topics is needed for success in college.” All items were 6-choice Likert statements with about half positively worded and about an equal number for each of the two attitudinal constructs under consideration.

Preparation of Teacher Profile Questionnaire

In Fishbein and Ajzen's model, behavior is a function of intended behaviors, which are a function of attitudes toward the behavior. Thus, the Teacher Profile Questionnaire was developed to identify the actual behaviors of teachers concerning the teaching of probability and statistics topics. Items selected for the Teacher Profile Questionnaire were designed to ascertain *coverage* of stochastics. Teachers were asked questions concerning the statistical topics that are part of their curriculum and those that are omitted.

Other demographical items were also on the questionnaire. The educational level, years of teaching experience, and experiences and education in mathematics and statistics were solicited from the respondents. These factors were analyzed for significance in the final study.

Administering the Pilot Scale

Selecting the Sample

A sample of secondary school mathematics teachers in the target population

was solicited to complete the preliminary TATSP Scale. It was not feasible to select a random sample. Instead, the cluster sampling technique was used. Several mathematics departments of public high schools in East Tennessee were contacted and asked to participate. From the group of schools that agreed to participate, five were randomly selected. Then, every math teacher at the chosen schools was asked to complete the preliminary scale.

Data Collection for the Pilot Scale

Copies of the scale were sent to the mathematics department chairpersons at the five schools selected to participate in the pilot study. The chairpersons distributed the scale to the math teachers at their respective schools. The scales were distributed with a cover letter, Appendix A, which explained to the teacher that completion of the survey was voluntary and insured confidentiality of the data collected. Separate self-addressed stamped envelopes were provided for the return of the completed scale. Of the 46 scales distributed 32 were completed and returned for a rate of 70 percent. Appendix B presents the pilot TATSP scale and the Teacher Profile Questionnaire.

Preparation of the Final Scale

Analysis of Data from Pilot Scale

About the Sample

The mathematics teachers who responded to the pilot scale had a wide range of teaching experience and knowledge of statistics. The mean years of teaching experience was 15.6 with a standard deviation of 9.8. Twenty-one of 32 reported that they had taken at least three college courses with a moderate amount of statistical content, and seven indicated they covered statistical topics in their math classes at a moderate to high level.

Scoring the Pilot Scale

Each positively stated item was scored from 6 (strongly agree) to 1 (strongly disagree). The scoring of negatively stated items was reversed. The first 24 items solicited beliefs about the importance of teaching statistical topics, and the remaining 19 items ascertained whether teachers liked teaching such topics. Item scores were summed to obtain an importance and a liking score for each respondent. The two subscale scores were combined to obtain a total scale score.

Distribution of responses.

Table 1 provides the frequency of responses and directionality for each item. There were 32 responses to each of the 43 items. The frequency distribution revealed a limited range of responses for some items, but there was no item for which all teachers responded identically. Table 2 shows the summary statistics for each item. Three items - 5, 12, and 15 - had mean scores less than a neutral 3.5 even after reversing for negatively worded items. Teachers on average said that statistical topics are important for most students later in life, especially college-bound students. They also enjoy teaching the topics and believe that their students enjoy statistics. However, they indicated that there is not enough time to cover statistical topics in regular high school math classes such as algebra.

The total scale scores ranged from 144 to 235. The mean score was 185 with a standard deviation of 20. The ranges for the importance and liking subscales were 51 and 56, respectively. Scores for teachers with between 10 and 20 years teaching experience were higher on both subscale scores and total score than teachers with either more or less experience, but the differences were not significant. Teachers who had taken more statistics in college had higher mean scores on all scales, but not significantly. Table 3 provides a listing of subscale scores, total scores, and reported coverage of statistical topics. Table 4 provides summary statistics for scale scores and coverage and a comparison of these scores by teaching experience and statistics education.

Table 1: Distribution of responses of items on 43-item pilot scale

Item No.	Direction	Response					
		1	2	3	4	5	6
1	positive	0	0	0	6	17	9
2	positive	0	0	2	10	17	3
3	negative	3	8	10	7	3	1
4	positive	0	0	0	11	15	6
5	negative	1	2	1	7	11	10
6	negative	2	12	7	8	2	1
7	positive	0	0	0	11	17	4
8	positive	0	1	1	10	17	3
9	positive	0	0	0	12	14	6
10	negative	6	9	9	5	2	1
11	negative	3	10	12	6	0	1
12	negative	2	5	5	12	7	1
13	negative	2	5	3	14	6	2
14	positive	0	0	1	7	18	6
15	positive	0	0	0	8	19	5
16	negative	9	12	3	7	1	0
17	positive	0	1	0	14	15	2
18	negative	3	11	9	7	0	2
19	positive	0	2	1	14	12	3
20	positive	0	4	5	11	11	1
21	negative	6	14	8	4	0	0
22	negative	6	17	6	1	2	0
23	positive	0	1	2	10	15	4

Table 1: (cont.)

Item No.	Direction	Response					
		1	2	3	4	5	6
24	positive	1	2	4	16	8	1
25	negative	4	13	8	5	2	0
26	positive	0	0	2	15	14	1
27	positive	0	2	6	14	9	1
28	negative	1	11	8	9	3	0
29	positive	0	2	2	8	17	3
30	negative	9	8	6	6	1	2
31	positive	0	1	2	10	16	3
32	positive	0	0	0	19	10	3
33	negative	2	11	9	7	3	0
34	positive	0	0	3	11	16	2
35	negative	2	7	13	9	0	1
36	positive	0	2	6	10	13	1
37	positive	0	1	9	12	9	1
38	negative	1	6	13	9	3	0
39	positive	0	1	2	14	13	2
40	positive	0	0	3	13	14	2
41	positive	0	0	4	15	11	2
42	negative	3	13	11	4	1	0
43	positive	0	1	2	9	14	6

Table 2: Summary statistics for 43-Item pilot scale after reversing scores of negative items

Item	Mean	Std Dev
1	5.09	0.69
2	4.66	0.75
3	3.94	1.24
4	4.84	0.72
5	2.28	1.30
6	4.03	1.20
7	4.78	0.66
8	4.63	0.83
9	4.81	0.74
10	4.28	1.30
11	4.22	1.07
12	3.28	1.26
13	3.28	1.30
14	4.91	0.73
15	4.91	0.64
16	4.66	1.21
17	4.53	0.76
18	4.13	1.24
19	4.41	0.95
20	4.00	1.08
21	4.69	0.93
22	4.75	1.02
23	4.59	0.91
24	3.97	1.03

Table 2: (cont.)

Item	Mean	Std Dev
25	4.38	1.10
26	4.44	0.67
27	4.03	0.93
28	3.94	1.08
29	4.53	0.98
30	4.38	1.48
31	4.56	0.88
32	4.50	0.67
33	4.06	1.11
34	4.53	0.76
35	3.97	1.03
36	4.16	0.99
37	4.00	0.92
38	3.78	0.97
39	4.41	0.84
40	4.47	0.76
41	4.34	0.79
42	4.41	0.95
43	4.69	0.97

Note. $n = 32$ for each item.

Table 3: Total score, subscale scores, and coverage ranked by total score for pilot scale

Importance	Liking	Total	Coverage
128	107	235	4
114	112	226	4
130	88	218	2
108	104	212	4
130	77	207	4
112	85	197	2
94	103	197	4
92	104	197	3
99	97	196	3
108	85	193	2
113	78	191	4
101	89	190	4
110	79	189	4
104	85	189	5
105	82	187	3
99	87	186	4
104	81	185	4
96	89	185	3
109	71	180	4
102	78	180	4
102	75	177	4
95	81	176	4
89	85	174	4
94	79	173	2

Table 3: (cont.)

Importance	Liking	Total	Coverage
104	68	172	3
96	76	172	4
101	64	165	2
94	70	164	4
79	83	162	4
93	68	161	3
96	60	156	4
88	56	144	3

Table 4: Comparison of coverage, subscale, total scores for pilot scale by teaching experience and statistics education

	Overall	by Teaching Experience ^a			by Stat Education ^b	
		0-9	10-19	20+	1-2	3+
Import:						
Mean	103	103	105	103	101	104
Std Dev	11.41	10.53	12.01	12.59	8.94	12.59
<i>n</i>	32	11	9	12	11	21
Liking:						
Mean	82	81	89	78	75*	86*
Std Dev	12.67	14.57	13.00	8.72	10.31	12.54
<i>n</i>	32	11	9	12	11	21
Total:						
Mean	185	184	194	180	177	190
Std Dev	20.03	23.07	20.30	16.10	16.76	20.43
<i>n</i>	32	11	9	12	11	21
Coverage:						
Mean	3.50	3.64	3.44	3.42	3.27	3.62
Std Dev	.80	0.50	0.73	1.08	0.90	0.74
<i>n</i>	32	11	9	12	11	21

^aComparison of means by analysis of variance procedure.

^bComparison of means by two-tail *t* tests.

* $p < 0.05$

Scoring of Personal Profile Questionnaire

Most teachers indicated that they had taken some courses in college that covered statistics. In fact, nearly three fourths had taken three or more courses. However, only one fourth reported that they covered any statistical topics thoroughly in their classes. If they had complete control of the curriculum, over half of the teachers said they would thoroughly cover some statistical topics. In scoring the responses to items on the questionnaire, numerical values were assigned to ordinal responses. For example, coverage was scores from none (1) to thoroughly (5). Table 5 shows the numerical scoring criteria as well as the frequency of responses to each item.

Correlation of Scale Scores and Questionnaire Variables

Pearson product moment r correlation coefficients were computed for the correlation between total scores and subscale scores and tested for significance using one-tail tests since positive correlations were hypothesized. Total score was strongly correlated with importance ($r = 0.81$) and liking ($r = 0.85$). However, the correlation between the two subscales was a modest $r = 0.38$, indicating that the two scales are measuring different constructs.

Correlations were also computed between scale scores and questionnaire variables. Total score and the liking subscale score were significantly correlated with the number of statistics courses taken.

Table 5: Numerical scoring and distribution of responses for items on the Personal Profile Questionnaire used with the pilot scale

Item	Responses				
Teaching Experience	0-9 years		10-19 years		20+ years
Numerical Score	0		1		2
Number of Responses	11		9		12
Statistics Courses	One	Two	Three	Four	Five
Numerical Score	0	0	1	1	1
Number of responses	5	6	8	8	5
Actual Coverage	None	Few	Light	Modest	Heavy
Numerical Score	1	2	3	4	5
Number of Responses	0	9	12	10	1
Intended Coverage	None	Few	Light	Modest	Heavy
Numerical Score	1	2	3	4	5
Number of Responses	0	5	7	19	1

No other significant correlations were found between scale scores and variables on the questionnaire, including coverage of statistical topics. Among the questionnaire variables, the number of statistics courses taken in college was negatively correlated with the years of teaching experience. Correlation coefficients are presented in Table 6.

Estimates of Internal Consistency

Cronbach's coefficient alpha estimates of internal consistency were computed for both subscales and for the overall scale. These ranged from 0.85 to 0.94. The coefficients are listed in Table 7.

Selection of Items for the Final Scale

The main purpose of the pilot study was to determine which items to discard or modify for the final scale. This process involved two steps. First conceptually valid items which reflect the emotional intensity of attitude were selected. Next, from those remaining, homogenous items were selected.

For an item to be conceptually valid, it must have responses in both a favorable and unfavorable direction by the respondents. Also, it should discriminate between those subjects with positive attitudes and those with negative ones.

Several items were discarded because they had response distributions that clustered at one end. For instance, items 1, 2, 4, 7, 9, and 15 were highly

Table 6: Correlations of scale scores and questionnaire variables for the 43-item pilot scale

	Total	Import	Liking	Teaching Experience	Statistics Courses	Actual cover	Intended Cover
Total	1.00						
Import	0.81*	1.00					
Liking	0.85*	0.38	1.00				
Teach exp.	-0.14	-0.05	-0.17	1.00			
Stat Course	0.37*	0.18	0.42*	-0.34*	1.00		
Actual Cover	0.25	0.16	0.25	0.08	0.32	1.00	
Intend Cover	0.05	-0.07	0.14	-0.17	0.24	0.37*	1.00

Note. All tests one-tail, $n = 32$.

* $p < 0.05$

Table 7: Estimates of internal consistency as measured by Cronbach's coefficient alpha for 43-item pilot scale and subscales

Scale/Subscale	Coefficient Alpha
Total	0.92
Importance	0.85
Liking	0.94

skewed in the "agree" direction. Thus, they were not included in the final scale.

To determine item discriminating quality, the teachers were split into high and low groups for each subscale. For each construct two extreme groups were established by identifying the subjects with the highest 25% of scores for that construct to form the high group and the subjects with the lowest 25% to form the low group. For each item the mean score of the high group was compared to the mean score of the low group by the two-sample t-test for independent samples. Items with statistically significant different means were considered to have discriminating quality.

The second step in selecting items was to choose items which measure the attitude under investigation. The Pearson product moment r correlation coefficients between each item and the appropriate subscale were computed and tested for significance using one-tail tests since positive correlations were expected.

Table 8 presents the results of the item analysis. Eleven of the 24 importance subscale items were selected for the final scale. Thirteen items were chosen for the liking subscale. Preference was given to a short scale that maintained satisfactory internal consistency. For the 24 item reduced scale, Cronbach's coefficient alpha estimates of internal consistency were 0.85, 0.91, and 0.89 for the importance subscale, the liking subscale, and the total scale, respectively.

Table 8: Correlation of item and subscale scores and discriminating quality of items on 43-item pilot scale

Item Number	Item-Subscale Correlation	Discriminating Quality	Selected Final Scale
I1	0.22	2.00	
I2	0.54*	2.53*	
I3	0.64*	3.82*	YES
I4	0.62*	4.00*	
I5	-0.02	0.76	
I6	0.16	0.43	
I7	0.46*	2.50*	
I8	0.41*	1.18	
I9	0.53*	3.02*	
I10	0.50*	2.74*	YES
I11	0.31	1.52	
I12	0.48*	2.63*	YES
I13	0.64*	2.45*	YES
I14	0.52*	2.11*	
I15	0.67*	4.02*	
I16	0.42	2.09	
I17	0.46*	3.41*	
I18	0.68*	4.06*	YES
I19	0.57*	4.40*	YES
I20	0.60*	3.96*	YES
I21	0.63*	4.27*	YES
I22	0.49*	4.47*	YES
I23	0.68*	4.47*	YES

Table 8: (cont.)

Item Number	Item-Subscale Correlation	Discriminating Quality	Selected Final Scale
I24	0.71*	3.42*	YES
L25	0.68*	5.43*	YES
L26	0.81*	7.24*	
L27	0.82*	5.41*	YES
L28	0.77*	8.00*	YES
L29	0.41	2.71	
L30	0.71*	4.40*	YES
L31	0.66*	3.21*	YES
L32	0.59*	3.75*	
L33	0.67*	4.00*	YES
L34	0.61*	3.00*	
L35	0.55*	2.75*	YES
L36	0.55*	4.13*	YES
L37	0.75*	3.96*	YES
L38	0.63*	2.79*	YES
L39	0.53*	3.53*	YES
L40	0.67*	3.00*	
L41	0.69*	3.24*	
L42	0.64*	5.37*	YES
L43	0.71*	4.82*	YES

Note: All tests one-tail, $n = 32$.

* $p < 0.05$

Administering the Final Scale

Selecting the Sample

A sample of secondary school mathematics teachers in the target population was solicited to complete the final TATSP Scale. It was not feasible to select a completely random sample. Instead, the cluster sampling technique was used. Each of the mathematics departments of public high schools in East Tennessee that supported the pilot study agreed to participate in the final study. From those schools, a random sample of 11 was chosen. Then, every math teacher at the chosen schools was asked to complete the TATSP scale.

Data Collection for the Final Scale

Copies of the scale were sent to the mathematics department chairpersons at the 11 schools selected to participate in the final study. The chairpersons distributed the scale to the math teachers at their respective schools. As with the pilot scale, the final scales were distributed with a cover letter, Appendix C, which explained to the teacher that completion of the survey was voluntary and insured confidentiality of the data collected. Separate self-addressed stamped envelopes were provided for the return of the completed scale. Of the 90 scales distributed 58 were completed and returned for a rate of 64 percent. Appendix D presents the final TATSP scale and Teacher Profile Questionnaire.

CHAPTER 5

RESULTS

Analysis of Data from Final Scale

About the Sample

The sample consisted of 58 mathematics teachers from 11 public high schools in East Tennessee. The mathematics teachers who responded to the final scale had a wide range of teaching experience and statistics education. The mean years of teaching experience was 14.4 with a range from one to 33 years. Two-thirds of the 58 respondents reported that they had taken at least three college courses with a moderate amount of statistical content.

Thirty-two of the 58 teachers who completed the final scale also participated in the pilot study. To determine whether participation in testing of the pilot scale may have affected teachers' scores on the final scale, total and subscale scores were compared for those who had earlier completed the pilot scale and those who had not. No significant differences were found in mean scores for any of the scales.

Distribution of Responses

The scoring for the final scale was the same as in the pilot scale. Each

positively stated item was scored from 6 (strongly agree) to 1 (strongly disagree). The scoring of negatively stated items was reversed. The first 11 items solicited beliefs about the importance of teaching statistical topics, and the remaining 13 items ascertained whether teachers liked teaching such topics. Item scores were summed to obtain an importance and a liking score for each respondent. The two subscale scores were combined to obtain a total scale score.

Table 9 provides the frequency of responses and directionality for each item. There were 58 responses to each of the 24 items. For nearly all items, the frequency distribution revealed a wide range of responses. Table 10 shows the summary statistics for each item after reversing scores for negative items. Item four, "Knowledge of statistical topics is not needed by students," was the only item in which the mean was less than a neutral 3.5. The standard deviations were greater than 1.00 for all items except two - Q22 and Q24.

Teachers on average said that constructing and reading tables and graphs were the were important skills for students later in life. They also indicated that they enjoyed teaching those skills. The respondents slightly agreed that statistical topics are important for most students later in life, especially college-bound students.

The total scale scores ranged from 58 to 144. The mean score was 101 with a standard deviation of 17. The ranges for the importance and liking subscales were 43 and 49, respectively. Scores for teachers with less than 10 years

Table 9: Distribution of responses of items on 24-item final TATSP scale

Item No.	Direction	Response					
		1	2	3	4	5	6
1	negative	7	16	15	12	7	1
2	negative	10	18	16	5	6	3
3	negative	3	13	15	15	10	2
4	negative	5	10	9	19	12	3
5	negative	9	14	17	14	2	2
6	positive	1	3	4	19	23	8
7	positive	0	6	9	21	18	4
8	negative	17	15	18	6	1	1
9	negative	21	20	12	2	3	0
10	positive	0	3	3	17	25	10
11	positive	1	3	9	25	16	4
12	negative	7	20	15	11	2	3
13	positive	2	5	9	23	15	4
14	negative	5	17	18	11	6	1
15	negative	14	18	10	12	2	2
16	positive	1	2	4	15	28	8
17	negative	8	19	14	10	7	0
18	negative	9	16	19	11	3	0
19	positive	0	4	10	19	19	6
20	positive	2	2	15	21	13	5
21	negative	4	8	25	15	6	0
22	positive	0	3	3	22	25	5
23	negative	11	18	19	7	3	0
24	positive	0	1	5	19	22	11

Table 10: Summary statistics for 24-Item Final TATSP Scale after reversing scores of negative items

Item	Mean	Std Dev
1	4.02	1.28
2	4.21	1.40
3	3.62	1.25
4	3.45	1.37
5	4.14	1.25
6	4.45	1.10
7	4.09	1.08
8	4.66	1.16
9	4.93	1.09
10	4.62	1.01
11	4.10	1.04
12	4.17	1.27
13	3.97	1.17
14	4.02	1.19
15	4.41	1.34
16	4.57	1.04
17	4.19	1.23
18	4.29	1.11
19	4.22	1.08
20	3.97	1.14
21	3.81	1.03
22	4.45	0.92
23	4.47	1.10
24	4.64	0.95

Note: $n = 58$ for each item.

teaching experience were higher on both subscale scores and total score than teachers with either more experience, but the differences were not significant. Teachers who had taken more statistics in college had higher mean scores on all scales, significantly for the total scale and the liking subscale. Table 11 provides a listing of subscale scores, total scores, and reported coverage of statistical topics. Table 12 provides summary statistics for scale scores and a comparison of these scores by coverage group. Teachers in the high coverage group introduce many statistical topics and may cover some of them in depth. The low coverage group consists of teachers who skip all statistical topics or introduce only a few of them. Table 13 presents summary statistics for scale scores and coverage and a comparison of these scores by teaching experience and statistics education.

Scoring of Personal Profile Questionnaire

As in the pilot study, numerical values were assigned to ordinal responses in scoring the responses to items on the questionnaire. For example, coverage was scores from none (1) to thoroughly (5). Table 14 shows the numerical scoring criteria as well as the frequency of responses to each item.

On average, teachers had completed three college courses with a moderate amount of statistical content. They indicated that some statistical topics are introduced but few, if any, are thoroughly covered. However, most would increase their statistical coverage if they had more control of the curriculum.

Table 11: Total score, subscale scores, and coverage ranked by total score for the 24-item final TATSP Scale

Importance	Liking	Total	Coverage
66	78	144	4
61	78	139	4
59	73	132	3
61	70	131	4
50	76	126	4
53	70	123	3
50	72	122	4
56	66	122	4
56	65	121	2
61	60	121	3
55	64	119	4
64	52	116	2
47	63	110	5
46	63	109	2
49	60	109	3
43	66	109	4
53	55	108	5
44	63	107	3
53	54	107	3
49	56	105	3
49	56	105	4
51	54	105	2
48	56	104	4
44	59	103	3

Table 11: (cont.)

Importance	Liking	Total	Coverage
46	57	103	2
47	56	103	3
44	59	103	3
39	63	102	2
48	54	102	3
45	56	101	3
45	56	101	4
53	47	100	3
48	50	98	2
50	47	97	3
47	50	97	3
49	48	97	3
37	57	97	3
42	54	96	1
51	43	94	2
40	53	93	3
46	47	93	2
39	54	93	2
41	52	93	2
38	55	93	3
47	45	92	2
41	51	92	2
44	47	91	2
38	51	89	2

Table 11: (cont.)

Importance	Liking	Total	Coverage
39	49	88	3
49	37	86	3
40	45	85	2
38	46	84	3
29	52	81	3
28	50	78	2
39	37	76	1
23	37	60	2
29	29	58	2

Table 12: Comparison of total and subscale scores by coverage group

Scale/Subscale	Low Coverage	High Coverage
Total		
Mean	92*	107*
Std Dev	15.7	15.2
<i>n</i>	22	36
Importance		
Mean	43*	49*
Std Dev	9.1	7.7
<i>n</i>	22	36
Liking		
Mean	49*	59*
Std Dev	9.2	9.6
<i>n</i>	22	36

Note: Comparison of means by two-tail *t* tests.

* $p < 0.05$

Table 13: Comparison of coverage, subscale, total scores for 24-item final TATSP scale by teaching experience and statistics education

	Overall	by Teaching Experience ^a			by Stat Education ^b	
		0-9	10-19	20+	1-2	3+
Import						
Mean	46	48	45	45	44	48
Std Dev	8.7	8.2	10.0	8.0	7.0	9.2
<i>n</i>	58	22	16	20	19	39
Liking						
Mean	55	58	56	52	51*	57*
Std Dev	10.5	10.4	12.5	8.0	8.4	10.9
<i>n</i>	58	22	16	20	19	39
Total						
Mean	101	106	100	97	95*	105*
Std Dev	17.1	17.0	21.1	12.3	13.2	17.9
<i>n</i>	58	22	16	20	19	39
Cover						
Mean	2.8	3.1	2.8	2.7	2.4*	3.1*
Std Dev	0.89	0.97	0.86	0.81	0.96	0.79
<i>n</i>	58	22	16	20	19	39

^aComparison of means by analysis of variance procedure.

^bComparison of means by two-tail *t* tests.

* $p < 0.05$

Table 14: Numerical scoring and distribution of responses for items on the Personal Profile Questionnaire used with the final scale

Item	Responses				
Teaching Experience	0-9 years		10-19 years		20+ years
Numerical Score	0		1		2
Number of Responses	22		16		20
Statistics Courses	One	Two	Three	Four	Five
Numerical Score	0	0	1	1	1
Number of responses	11	8	17	11	11
Actual Coverage	None	Few	Light	Modest	Heavy
Numerical Score	1	2	3	4	5
Number of Responses	2	20	23	11	2
Intended Coverage	None	Few	Light	Modest	Heavy
Numerical Score	1	2	3	4	5
Number of Responses	2	11	23	19	3

Analysis of Data for Research Hypotheses

Statistical analyses were performed to test the research hypotheses and assist in answering the research question. The following analyses were conducted.

1. Pearson's product moment correlations were computed between each item that measures the importance dimension and the total importance score, and tested at the 0.05 level of significance. Similar tests were performed on items measuring the liking dimension and the total liking score.

2. The discriminating quality of each item was computed as in the pilot scale. For each construct two extreme groups was established by identifying the subjects with the highest 25% of scores for that construct to form the high group and the subjects with the lowest 25% to form the low group. For each item the mean score of the high group was compared to the mean score of the low group by the two-sample t-test for independent samples.

3. The Pearson product moment r correlation coefficient between the measure of the two attitude constructs, importance of teaching stochastics and the liking of teaching stochastics, and the total scale score, were computed and tested for significance at a 0.05 level of significance.

4. The Pearson product moment r correlation coefficients between the self-reported coverage involving teaching statistical topics and the scale scores will be computed and tested for significance at a 0.05 level.

5. The Pearson product moment r correlation coefficients between the scale

scores and items on the personal profile questionnaire were computed and tested for significance at a 0.05 level.

6. Multiple linear regression was performed with the self-reported coverage of statistical topics as the dependent variable and the two estimates of the attitude constructs as predictors. Regression coefficients was tested for statistical significance at the 0.05 level of significance. The final regression equation was used to generate predicted levels of coverage. The Pearson product moment correlation coefficient was computed between predicted level of coverage and the reported level and tested for significance.

Correlation of Variables

Correlation of scale scores and questionnaire variables.

As in the pilot study, Pearson product moment r correlation coefficients were computed for the correlation between total scores and subscale scores and tested for significance using one-tail tests since positive correlations were hypothesized. Total score was strongly correlated with the importance subscale score ($r = 0.86$) and the liking subscale score ($r = 0.91$). However, the correlation between the two subscales was not as strong at $r = 0.58$, giving some indication that the two scales are measuring different constructs.

Correlations were also computed between scale scores and questionnaire variables with several of them significant. The number of statistics courses

taken correlated significantly with the total score and the liking subscale score. The reported coverage of statistics was significantly correlated with total score and both subscale scores. Years teaching experience was slightly correlated with the liking subscale score, but not significantly correlated with any of the other scale scores.

Among the questionnaire variables, the number of statistics courses taken in college was positively correlated with coverage in the classroom. Of course, intentions concerning coverage of statistical topics was highly correlated with the actual reported coverage. Correlation coefficients are presented in Table 15.

Item Analysis

Item-subscale and item-coverage correlations and discriminating quality for each item were computed to verify contributions to the scale. Results are presented in Table 16. All 24 items were able to discriminate at a 0.05 significance level and were also significantly correlated with its corresponding subscale. All but five of the items were significantly correlated with the reported coverage of statistical topics.

Table 15: Correlations of scale scores and questionnaire variables for the 24-item final TATSP scale

	Total	Import	Liking	Teaching Experience	Statistics Courses	Actual cover	Intended Cover
Total	1.00						
Import	0.86*	1.00					
Liking	0.91*	0.58*	1.00				
Teach Exp.	-0.24	-0.15	-0.27*	1.00			
Stat Course	0.28*	0.20	0.28*	-0.22*	1.00		
Actual Cover	0.55*	0.41*	0.56*	-0.16	0.32*	1.00	
Intend Cover	0.49*	0.45*	0.42*	-0.07	0.29*	0.74*	1.00

Note. All tests one-tail, $n = 58$.

* $p < 0.05$.

Table 16: Item-subscale and item-coverage correlations and discriminating quality of items on 24-item final TATSP scale

Item Number	Item-Subscale Correlation	Discriminating Quality	Item-Coverage Correlation
I1	0.66*	5.58*	0.19
I2	0.65*	6.24*	0.22
I3	0.72*	8.32*	0.21
I4	0.70*	5.45*	0.12
I5	0.68*	5.13*	0.35*
I6	0.67*	5.47*	0.38*
I7	0.69*	5.08*	0.38*
I8	0.54*	5.71*	0.29*
I9	0.62*	4.90*	0.37*
I10	0.68*	4.07*	0.23
I11	0.75*	6.26*	0.38*
L12	0.76*	5.79*	0.35*
L13	0.81*	6.78*	0.45*
L14	0.69*	5.12*	0.37*
L15	0.71*	6.54*	0.42*
L16	0.67*	5.06*	0.28*
L17	0.66*	5.05*	0.58*
L18	0.64*	6.55*	0.35*
L19	0.72*	6.22*	0.51*
L20	0.78*	7.54*	0.37*
L21	0.75*	5.20*	0.46*
L22	0.63*	5.11*	0.34*
L23	0.79*	9.93*	0.45*
L24	0.75*	6.81*	0.28*

Note: All tests one-tail, $n = 58$.

* $p < 0.05$

Predicting Usage and Discriminating Between Low and High Users

Multiple linear regression was performed with the self-reported coverage of statistical topics as the dependent variable and the two subscale scores as predictors. At the 0.05 significance level, this resulted in a prediction equation based solely on the liking subscale score. However, multicollinearity existed in the model. Therefore, principal component regression was performed which resulted in a weighted average of subscale scores for predicting coverage.

WTAVG = 0.2677 I + 0.8378 L, where I = Important score and L = Liking score.

The correlation of coverage with the derived weighted average was 0.57, slightly higher than the 0.55 correlation between coverage and total scale score.

The 58 respondents were split into high and low coverage groups based on the self-reported level of coverage. The high coverage group consisted of those teachers who introduce many statistical topics and may cover some of them in depth. Thirty-six of the respondents fell into this group. Discriminant analysis was performed to predict coverage group. The total score and the weighted average score were used in separate models. Identical results were obtained for both models; 52 of 58 or 72 percent of the teachers were correctly classified. For both groups, about 72 percent of the teachers were correctly classified.

Reliability

Estimates of Internal Consistency

Cronbach's coefficient alpha estimates of internal consistency were computed for both subscales and for the overall scale. The Liking subscale had the highest measure at 0.93. The coefficient for the Important subscale was 0.85, and for the total scale 0.92.

Test-Retest Reliability

Thirty-two of the 58 respondents to the final scale also completed the pilot one. For their pilot scale scores the responses to the 19 items not included on the final TATSP scale were ignored and their scales rescored. Test-retest correlation coefficients were computed between the reduced scale and final scale scores. The interval between completion of the pilot and final scales was about four weeks. The coefficients are presented in Table 17.

Table 17: Test-Retest correlation coefficients

Scale/Subscale	Coefficient
Total	0.82*
Import	0.74*
Liking	0.79*

Note: All tests single-tail, n=32.

* $p < 0.05$

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

Description of the TATSP Scale

In constructing the TATSP scale, two major steps were carried out. First a pilot scale consisting of 43 items was constructed to solicit evaluative beliefs about either the importance or liking of teaching statistical topics. Next, the pilot scale was completed by 32 mathematics teachers. From this data, an item was selected for the final scale based on its item-subscale correlations and discriminating quality. This resulted in a 24-item final TATSP scale that was valid and homogeneous. It consisted of two reliable and unidimensional subscales: (1) importance of teaching statistical topics and (2) liking of teaching statistical topics. The final scale was then completed by 58 teachers.

Reliability

Cronbach's coefficient alpha and test-retest correlations indicate that the TATSP scale and both subscales are reliable. The internal consistency of the scales range from 0.85 for the importance subscale to 0.94 for the liking one. Coefficient alpha for the entire scale is 0.92. Test-retest correlations are all strong.

The data also indicates that the 24 items on the final scale are homogenous

within each subscale. Correlations between each item and its subscale are significant. High coefficient alphas for both subscales also suggest homogeneity.

Validity

Construct Validity

Construct validity requires evidence that the measurements provided by the TATSP scale be consistent with Fishbein's theory of reasoned action. In other words, if the TATSP is a measure of the stated attitudes, then the scale scores should be related to the actual behaviors of the teachers concerning teaching statistics.

Support for construct validity is provided by the known-group approach to validation. Teachers' level of interest in statistics is determined by the self-reported level of coverage of statistical topics. The mean score on the TATSP scale of teachers who show high interest in statistics are higher than the mean score of teachers who reject teaching statistics.

The moderate correlation between the two subscales (0.58) also provide support of construct validity. The evidence is that two different constructs are being measured. The combination of the two subscales provides a better predictor of coverage of statistics than either subscale alone. A weighted average of components is more strongly related to coverage than simply

combining the two subscale scores, but the difference does not seem to be important.

Some additional evidence for construct validity includes the item-subscale correlations and discriminating quality of the items. For the final data most items on the final scale are at least moderately related to the subscale score, are correlated with usage, and do discriminate.

Criterion-Related Validity

Criterion-related validity requires that the attitude scores be correlated with scores on another variable, usually behavior. The TATSP is valid for classification of mathematics teachers into high and low groups regarding the amount of coverage statistical topics receive in their classrooms. Regression models provide predictions of coverage which correlate with the self-reported coverage around 0.57. About 72 percent of the teachers are correctly classified as members of the low or high coverage groups using discriminant analysis. Nearly identical results are obtained using models based only on total scores or on the derived weighted average to classify teachers.

Summary of Reliability and Validity

A major purpose of this study was to construct the Teachers' Attitude Toward Statistics and Probability (TATSP) scale by using Fishbein's theory of reasoned action. The primary hypothesis was that Fishbein's theory is a valid

model when the object under consideration is teaching statistics. In other words, the relationship between a teacher's beliefs about teaching statistics, attitudes about teaching statistics, and behaviors involving teaching statistics, would be consistent with Fishbein's model. This model of behavior provided a sound basis for the development of the scale.

The scale that was developed presumed to measure beliefs concerning teaching statistical topics along two constructs: the importance and liking of teaching such topics. It was shown that the TATSP scale is valid and reliable when used for measuring mathematics teachers' attitudes toward teaching statistics.

Factors Associated with Coverage of Statistics

It was also shown that teachers who like statistics cover more statistical topics in their classrooms. Much of the literature asserted that the perceived importance of a topic is the main reason a topic is covered or class or omitted. However, the liking subscale score was much more highly correlated with coverage ($r = 0.56$) than the importance subscale ($r = 0.41$). Also, the liking score was a better predictor of coverage and was more successful at discriminating between high and low coverage teachers.

Secondary mathematics teachers who are more knowledgeable of statistics, measured by the number of college statistics courses completed, liked statistics more. Consequently, they tended to cover statistical topics more. The study

also showed that teachers who graduated from college recently had completed more statistics courses than their more experienced counterparts.

Conclusions

1. The TATSP Scale is valid and reliable when used for measuring the attitudes of secondary school mathematics teachers toward their coverage of statistical topics.

2. Fishbein's and Ajzen's theory of reasoned action and the general principles for constructing affective scales provide a sound basis for development of the scale.

3. Secondary mathematics teachers who are more knowledgeable of statistics like statistics more than their less experienced colleagues.

4. Secondary mathematics teachers who like to teaching statistical topics teach more statistical topics than their colleagues.

5. Secondary mathematics teachers who hold more positive attitudes toward teaching statistical concepts cover more statistical topics than their colleagues.

Recommendations

Suggestions for Further Study

This study should be replicated with a larger sample and should include

teachers from other geographical areas. The TATSP scale seems to be satisfactory, but should be used with a wide variety of teachers to determine if it functions well with other populations.

Validation of the scale needs to be examined further. Alternative methods of measuring teachers' attitudes toward teaching statistics, such as interviewing or observational techniques, should be developed. The convergent validity of the TATSP scale with these new measures would help determine the construct validity of the instruments.

Studies Using the TATSP Scale

The finding of differences in scores for teachers who had more knowledge of statistics may provide a method to change attitudes. Using the TATSP Scale as a pre-test and post-test, studies should be conducted on the effects of increasing teachers' knowledge of statistics on their attitudes toward teaching statistics. The effects of treatments such as in service training or even a college course on teachers' attitudes could be investigated.

The relationship between teachers' attitudes toward teaching statistics and their students' attitudes toward statistics needs to be examined. Many scales have been developed that measure students' attitudes toward statistics. For instance, Roberts and Bilderback's (1980) Statistics Attitude Survey or Wise's (1985) Attitudes Toward Statistics scale. The TATSP Scale could be used to measure the teachers' attitudes, and the relationship between the two

investigated. Many studies have looked at the relationship with other subject areas, but none have focused on statistics.

These are additional suggestions of studies that use the TATSP Scale:

1. The relationship between teachers' attitudes toward teaching statistics and their students' statistical achievement needs to be investigated.
2. The attitudes of teachers on the block schedule versus those on a traditional schedule should be studied to determine if there is a difference
3. The attitudes of teachers from different grade levels should be studied. The elementary school curriculum includes many statistical topics. A look at the attitudes of elementary teachers and a comparison of those attitudes to secondary teachers' attitudes would be of interest.

REFERENCES

- Aiken, L. R., Jr. (1970). Attitudes toward mathematics. Review of Educational Research, 40, 551-596.
- Aiken, L. R., Jr. (1974). Two scales of attitude toward mathematics. Journal of Research in Mathematics Education, 5, 67-71.
- Aiken, L. R., Jr. (1980). Attitude measurement and research. New Directions for Testing and Measurement, 7, 1-24.
- Anderson, L. W. (1981). Assessing affective characteristics in the schools. Boston: Allyn and Bacon.
- Auzmendi, E. (1991). Factors related to attitudes toward statistics: A study with a Spanish sample. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, 3-7 April. (ERIC Document Reproduction Service No. ED 333 049)
- Barr, R. (1988). Mathematics content. The Elementary School Journal, 88, 387-411.
- Benninga, J. S., Guskey, T. R., & Thornburg, K. R. (1981). The relationship between teacher attitudes and student perceptions of classroom climate. The Elementary School Journal, 82(1), 66-75.
- Bohrnstedt, G. W. (1970). Reliability and validity assessment in attitude measurement. In G. F. Summers (Ed.), Attitude Measurement. Chicago: Rand McNally.
- Brophy, J. E. (1986). Teaching and learning mathematics: Where research should be going. Journal for Research in Mathematics Education, 17, 323-346.
- Brown, C. A., & Baird, J. (1993). Inside the teacher: Knowledge, Beliefs, and attitudes. In P. S. Wilson (Ed.), Research ideas for the classroom: High school mathematics (pp. 245-259). New York: Macmillan.
- Buchanon, N. K. (1987). Factors contributing to mathematical problem-solving performance and exploratory study. Educational Studies in Mathematics, 18, 399-415.
- Burton, Hopkins, Johnson, Kaplan, Kennedy, & Schultz (1992). Mathematics plus. Orlando, FL: Harcourt Brace Jovanovich.

- Campbell, D. T., & Fiske, D. W. (1959). Convergent and divergent validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.
- Crano, W. D., & Brewer, M. B. (1973). Principles of research in social psychology. New York: McGraw-Hill.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.
- Confrey, J. (1986). A critique of teacher effectiveness research in mathematics education. Journal for research in Mathematics Education, 17, 347-360.
- Doob, L. W. (1947). The behavior of attitudes. Psychological Review, 54, 135-156.
- Dreeben, R. (1985). The social organization of mathematics and reading instruction. Chicago: American Educational Research Association.
- Ferguson, G. A. (1971). Statistical analysis in psychology & education. New York: McGraw-Hill.
- Fishbein, M. (Ed.) (1967). Readings in attitude theory and measurement. New York: McGraw-Hill.
- Fishbein, M., & Ajzen, I. (1975). Belief, attitude, intention and behavior: An introduction to theory and research. Reading, MA: Addison-Wesley.
- Fishbein, M., & Ajzen, I. (1980). Understanding attitudes and predicting social behavior. Englewood Cliffs, NJ: Prentice Hall.
- Frary, R., & Ling, J. (1983). A factor analytic study of mathematics anxiety. Educational and Psychological Measurement, 43, 985-993.
- Garrett, H. E. (1966). Statistics in psychology and education (6th ed.). New York: David McKay Company.
- Grey, D. R. (Ed.). (1983). Proceedings of the first international conference on teaching statistics. (ERIC Document Reproduction Service No. ED 262 977)
- Katz, D., & Scotland, E. (1959). A preliminary statement to a theory of attitude structure and change. In S. Koch (Ed.), Psychology: A study of a science. New York: McGraw-Hill.

- Krech, D., & Crutchfield, R. S. (1948). Theory and problems in social psychology. New York: McGraw-Hill.
- Lampert, M. (1985). How do teachers manage to teach? Perspectives on problems in practice. Harvard Educational Review, 55(2), 178-194.
- Larson, R., Kanold, T., & Stiff, L. (1993). Algebra 1. Lexington, MA: D. C. Heath.
- Lemon, N. (1973). Attitudes and their measurement. New York: John Wiley & Sons.
- Magnussen, D. (1967). Test theory (H. Mabon, Trans.). Reading, PA: Addison Wesley.
- Manre, Moser, Lobato, & Morror (1992). Mathematic corrections. Lexington, MA: D. C. Heath.
- Mendoza, L. R., & Swift, J. (1981). Why teach statistics and probability-a rationale. In A. P. Shulte (Ed.), Teaching statistics and probability (pp. 1-7). Reston, VA: NCTM.
- Murphy, G., & Likert, R. (1937). Public opinion and the individual. New York: Harper.
- Murphy, K. R., & Davidshofer, C. O. (1991). Psychological testing: Principles and applications (2nd ed.). Boston: PWS Kent.
- National Council of Teachers of Mathematics. (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: Author.
- Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.
- Oppenheim, A. N. (1966). Questionnaire design and attitude measurement. New York: Basic Books.
- Osgood, C.E., Suci, G. J., & Tannenbaum, P. H. (1971). The measurement of meaning. Urbana, IL: University of Illinois Press.
- Porter, A., Floden, R., Freeman, D., Schmidt, W., & Schwille, J. (1988). Content determinants in elementary school mathematics. In D. Grouws & T. Cooney (Eds.), Perspectives on research on effective mathematics teaching (pp. 96-113). Reston, VA: National Council of Teachers of Mathematics.

- Quilter, D., & Harper, E. (1988). Why we didn't like mathematics, and why we can't do it. Educational Research, 30, 121-134.
- Roberts, D. M., & Bilderback, E. W. (1980). Reliability and validity of a statistics attitude survey. Educational and Psychological Measurement, 40, 235-238.
- Roberts, D. M., & Reese, C. M. (1987). A comparison of two scales measuring attitudes towards statistics. Educational and Psychological Measurement, 47, 759-764.
- Roberts, D. M., & Saxe, J. E. (1982). Validity of a statistics attitude survey: A follow-up study. Educational and Psychological Measurement, 42, 907-912.
- Satarso, T. (1992). Students' attitudes toward statistics. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Knoxville, TN, 11-13 November. (ERIC Document Reproduction Service No. ED 353 334)
- Schunk, D. (1985). Self-efficacy and classroom learning. Psychology in the Schools, 22, 208-223.
- Schwille, J., Porter, A., Belli, G., Floden, R., Freeman, D., Knappen, L., Kuhs, T., & Schmidt, W. (1982). Teachers as policy brokers in the content of elementary school mathematics. In L. Shulman & G. Sykes (Eds.), Handbook of teaching and policy (pp. 370-391). New York: Longman.
- Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning (pp. 465-494). New York: MacMillan.
- Shaughnessy, J. M., & Bergman, B. (1993). Thinking about uncertainty: Probability and Statistics. In P. S. Wilson (Ed.), Research ideas for the classroom: High school mathematics (pp. 177-197). New York: Macmillan.
- Shaw, M. E., & Wright, J. M. (1967). Scales for the measurement of attitudes. New York: McGraw-Hill.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. Educational Researcher, 15(2), 4-14.
- Shulman, L. S. (1988). Knowledge growth in teaching: A final report to the Spencer Foundation. Stanford, CA: Stanford University.

Shulte, A. P. (Ed.). (1981). Teaching statistics and probability (1981 Yearbook of the National Council of Teachers of Mathematics). Reston, VA: NCTM.

Stanley, J. C. (1971). Reliability. In R. L. Thorndike (Ed.), Educational measurement and evaluation. Englewood Cliffs, NJ: Prentice Hall.

Thurstone, L. L. (1931). The measurement of social attitudes. Journal of Abnormal and Social Psychology, 26, 249-269.

Turner, V. L. (1987). A theoretical framework for the study of attitudes toward mathematics (Doctoral dissertation, University of Oklahoma, 1987). Dissertation Abstracts International, 48,

Usiskin, Z. (1981). Decision-making in mathematics education. In L. Steen & D. Albers (Eds.), Teaching teachers, teaching students (pp. 43-55). Boston: Birkhauser.

Wise, S. L. (1985). The development and validation of a scale measuring attitudes toward statistics. Educational and Psychological Measurement, 45, 401-405.

Woehlke, P. L. (1991). An examination of the factor structure of Wise's "Attitude Toward Statistics" scale. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL, 3-7 April. (ERIC Document Reproduction Service No. ED 337 500)

APPENDICES

APPENDIX A
Cover Letter for Administration
of Pilot TATSP Scale

Dear Math Teacher:

I am conducting a study to develop and validate an instrument to measure mathematics teachers' attitudes toward teaching statistical topics in the normal high school mathematics curriculum. You are requested to complete a pilot version of that scale by responding to statements designed to measure your beliefs. You are also requested to complete a short section providing certain personal data. Your participation in this study will help the researcher develop and validate the scale.

There are no risks to you from participating. Your responses will be kept confidential and no record of your identity will be recorded except the name of your school; these will provide a means of matching data from the pilot and final versions of the scale to enable correlation of the responses. Only the investigator will have access to the forms with your responses and the data you provide. If you have any questions about the research, please contact Mike Winstead, 3237 Boomerang Lane, Knoxville, Tennessee, 37931 or call 423-691-7319.

Your participation in this study is voluntary, and you may refuse to participate without penalty. You may participate by responding to this pilot scale and later refuse, without penalty, to respond to the final version. By completing and returning this survey, you are indicating that you have read and understand the explanation of this study and agree to participate.

APPENDIX B

Pilot TATSP Scale

Mathematics Teachers' Coverage of Statistical Topics

Thank you for participating in this research. There are no correct or incorrect answers. Your responses are important and will be kept strictly confidential.

Please respond to every statement by circling the number which indicates the degree to which you agree or disagree with each statement.

	Strongly Disagree	Moderately Disagree	Slightly Disagree	Slightly Agree	Moderately Agree	Strongly Agree
	1	2	3	4	5	6
<i>1. Importance of teaching statistical topics</i>						
1. Statistical topics provide students with real-world application of mathematics.	1	2	3	4	5	6
2. Statistical topics should be taught because of their importance to students later in life.	1	2	3	4	5	6
3. Students don't need to understand statistical concepts to function in society.	1	2	3	4	5	6
4. An understanding of statistical topics is needed for many jobs.	1	2	3	4	5	6
5. There is not enough time to cover statistical topics in a course such as algebra.	1	2	3	4	5	6
6. Knowledge of statistical topics is not needed on college entrance exams.	1	2	3	4	5	6
7. Students must have the ability to read and interpret graphs to function efficiently in society.	1	2	3	4	5	6
8. Students should be taught how to construct graphs and tables from data.	1	2	3	4	5	6
9. It is important that students have experience solving real-life problems using statistics.	1	2	3	4	5	6
10. Knowledge of statistical concepts is not needed for success in college.	1	2	3	4	5	6
11. The content in high school courses such as Algebra I should not include topics from statistics.	1	2	3	4	5	6
12. Individuals do not need the ability to determine probabilities to function efficiently in society.	1	2	3	4	5	6

- | | | | | | | |
|---|---|---|---|---|---|---|
| 13. Knowledge of statistical concepts is not needed by everyone. | 1 | 2 | 3 | 4 | 5 | 6 |
| 14. Everyone needs to be able to recognize misleading statistics found in newspapers. | 1 | 2 | 3 | 4 | 5 | 6 |
| 15. Everyone needs to be able to interpret tables and graphs encountered daily. | 1 | 2 | 3 | 4 | 5 | 6 |
| 16. Calculating averages is a skill that students don't need for later success in school. | 1 | 2 | 3 | 4 | 5 | 6 |
| 17. Students should be able to recognize the appropriate uses of the different measures of central tendency (mean, median, and mode). | 1 | 2 | 3 | 4 | 5 | 6 |
| 18. Students don't need to be taught the different methods of computing an average of a set of data. | 1 | 2 | 3 | 4 | 5 | 6 |
| 19. Calculating simple probabilities is an important skill for students to learn. | 1 | 2 | 3 | 4 | 5 | 6 |
| 20. Probability concepts should be taught because they are needed for many jobs. | 1 | 2 | 3 | 4 | 5 | 6 |
| 21. I do not feel that statistical topics should be covered in the normal high school sequence. | 1 | 2 | 3 | 4 | 5 | 6 |
| 22. The ability to read graphs is not important to students later in life. | 1 | 2 | 3 | 4 | 5 | 6 |
| 23. It is important for students to be able to communicate using tables and graphs. | 1 | 2 | 3 | 4 | 5 | 6 |
| 24. Students need to understand the meaning of the variation of a set of data. | 1 | 2 | 3 | 4 | 5 | 6 |

II. Liking of teaching statistical topics

- | | | | | | | |
|---|---|---|---|---|---|---|
| 25. Statistical topics are boring to students. | 1 | 2 | 3 | 4 | 5 | 6 |
| 26. Teaching statistical topics allows my students to be actively involved in the learning process | 1 | 2 | 3 | 4 | 5 | 6 |
| 27. Students are easily motivated to learn statistical concepts. | 1 | 2 | 3 | 4 | 5 | 6 |
| 28. Constructing tables and graphs is uninteresting to students. | 1 | 2 | 3 | 4 | 5 | 6 |
| 29. Statistical topics are fun to teach. | 1 | 2 | 3 | 4 | 5 | 6 |
| 30. Statistical concepts are some of my least favorite topics to teach. | 1 | 2 | 3 | 4 | 5 | 6 |
| 31. Collecting and analyzing data is an interesting project for students. | 1 | 2 | 3 | 4 | 5 | 6 |
| 32. Statistical concepts provide interesting applications of some concepts discussed in regular math classes such as algebra. | 1 | 2 | 3 | 4 | 5 | 6 |

- | | | | | | | |
|---|---|---|---|---|---|---|
| 33. Calculating simple probabilities is an unenjoyable topic for students. | 1 | 2 | 3 | 4 | 5 | 6 |
| 34. I like to teach students how to calculate simple probabilities. | 1 | 2 | 3 | 4 | 5 | 6 |
| 35. I dislike teaching the concept of central tendency of a set of data. | 1 | 2 | 3 | 4 | 5 | 6 |
| 36. Constructing tables and graphs is enjoyable to students. | 1 | 2 | 3 | 4 | 5 | 6 |
| 37. Students are easily motivated to collect data. | 1 | 2 | 3 | 4 | 5 | 6 |
| 38. The different measures of central tendency are uninteresting topics to students. | 1 | 2 | 3 | 4 | 5 | 6 |
| 39. I enjoy teaching students how to develop tables and graphs. | 1 | 2 | 3 | 4 | 5 | 6 |
| 40. Teaching students to analyze tables is enjoyable. | 1 | 2 | 3 | 4 | 5 | 6 |
| 41. Learning to analyzing tables is an interesting lesson for students. | 1 | 2 | 3 | 4 | 5 | 6 |
| 42. Discussing real-life statistical applications is not interesting to students. | 1 | 2 | 3 | 4 | 5 | 6 |
| 43. Statistical topics are encouraging to students because they show real-life applications of mathematics. | 1 | 2 | 3 | 4 | 5 | 6 |

III. Personal Profile Questionnaire

44. Number of years teaching experience. _____
45. Number of college-level math courses with at least a moderate amount of coverage of statistical topics.
46. All teachers in Tennessee are to follow the State Curriculum Frameworks. Topics in probability and statistics are contained in the Frameworks for most high school mathematics courses. However, each teacher must make decisions about which topics to cover and the depth at which to cover them. During the past year, how would you describe the level of coverage in your classroom of statistical topics.
- A. Completely skipped all statistical topics.
 - B. Introduced a few of the topics.
 - C. Introduced many of the topics.
 - D. Covered some of the topics thoroughly.
 - E. Covered most of the topics thoroughly.
47. Suppose you are solely responsible for the choosing the content in a regular high school mathematics course such as prealgebra, algebra, geometry, or algebra II. What would be your intentions concerning coverage of the statistical topics that appear in the textbooks for those courses.
- A. Completely skip all statistical topics.
 - B. Introduce a few of the topics.
 - C. Introduce many of the topics.
 - D. Cover some of the topics thoroughly.
 - E. Cover most of the topics thoroughly.

APPENDIX C
Cover Letter for Administration
of Final TATSP Scale

Dear Math Teacher:

I am conducting a study to develop and validate an instrument to measure mathematics teachers' attitudes toward teaching statistical topics in the normal high school mathematics curriculum. You are requested to complete the final version of that scale by responding to statements designed to measure your beliefs. You are also requested to complete a short section providing certain personal data.

There are no risks to you from participating. Your responses will be kept confidential and no record of your identity will be recorded except your date of birth and the name of your school; these will provide a means of matching data from the pilot and final versions of the scale to enable correlation of the responses. Only the investigator will have access to the forms with your responses and the data you provide. If you have any questions about the research, please contact Mike Winstead, 3237 Boomerang Lane, Knoxville, Tennessee, 37931 or call 423-691-7319.

Your participation in this study is voluntary, and you may refuse to participate without penalty. Even if you participated in the pilot study, you may refuse, without penalty, to respond to the final version. By completing and returning this survey, you are indicating that you have read and understand the explanation of this study and agree to participate.

APPENDIX D

Final TATSP Scale

Mathematics Teachers' Coverage of Statistical Topics

Thank you for participating in this research. There are no correct or incorrect answers. Your responses are important and will be kept strictly confidential.

The statements are designed to elicit your beliefs about statistical topics and about teaching such topics in your math classes. Statistical topics are those typically focusing on collecting and analyzing data.

Please respond to every statement by circling the number which indicates the degree to which you agree or disagree with each statement.

Strongly Disagree	Moderately Disagree	Slightly Disagree	Slightly Agree	Moderately Agree	Strongly Agree
1	2	3	4	5	6
1. Students don't need to understand statistical concepts to function in society.				1	2 3 4 5 6
2. Knowledge of statistical concepts is not needed for success in college.				1	2 3 4 5 6
3. Individuals do not need the ability to determine probabilities to function efficiently in society.				1	2 3 4 5 6
4. Knowledge of statistical concepts is not needed by everyone.				1	2 3 4 5 6
5. Students don't need to be taught the different methods of computing an average of a set of data.				1	2 3 4 5 6
6. Calculating simple probabilities is an important skill for students to learn.				1	2 3 4 5 6
7. Probability concepts should be taught because they are needed for many jobs.				1	2 3 4 5 6
8. I do not feel that statistical topics should be covered in the normal high school sequence.				1	2 3 4 5 6
9. The ability to read graphs is not important to students later in life.				1	2 3 4 5 6
10. It is important for students to be able to communicate using tables and graphs.				1	2 3 4 5 6
11. Students need to understand the meaning of the variation of a set of data.				1	2 3 4 5 6
12. Statistical topics are boring to students.				1	2 3 4 5 6
13. Students are easily motivated to learn statistical concepts.				1	2 3 4 5 6
14. Constructing tables and graphs is uninteresting to students.				1	2 3 4 5 6
15. Statistical concepts are some of my least favorite topics to teach.				1	2 3 4 5 6

- | | | | | | | |
|---|---|---|---|---|---|---|
| 16. Collecting and analyzing data is an interesting project for students. | 1 | 2 | 3 | 4 | 5 | 6 |
| 17. Calculating simple probabilities is an unenjoyable topic for students. | 1 | 2 | 3 | 4 | 5 | 6 |
| 18. I dislike teaching the concept of central tendency of a set of data. | 1 | 2 | 3 | 4 | 5 | 6 |
| 19. Constructing tables and graphs is enjoyable to students. | 1 | 2 | 3 | 4 | 5 | 6 |
| 20. Students are easily motivated to collect data. | 1 | 2 | 3 | 4 | 5 | 6 |
| 21. The different measures of central tendency are uninteresting topics to students. | 1 | 2 | 3 | 4 | 5 | 6 |
| 22. I enjoy teaching students how to develop tables and graphs. | 1 | 2 | 3 | 4 | 5 | 6 |
| 23. Discussing real-life statistics applications is uninteresting to students. | 1 | 2 | 3 | 4 | 5 | 6 |
| 24. Statistical topics are encouraging to students because they show real-life applications of mathematics. | 1 | 2 | 3 | 4 | 5 | 6 |

II. Personal Profile Questionnaire

25. Years teaching experience: _____
26. Name of your school: _____
27. The number of college-level math courses I have taken with at least a moderate amount of coverage of statistical topics is about:
- zero one two three four five or more
28. All teachers in Tennessee are to follow the State Curriculum Frameworks. Topics in probability and statistics are contained in the Frameworks for most high school mathematics courses. However, each teacher must make decisions about which topics to cover and the depth at which cover them. During the past year, how would you describe the level of coverage in your classroom of statistical topics.
- A. Completely skipped all statistical topics.
 - B. Introduced a few of the topics.
 - C. Introduced many of the topics.
 - D. Covered some of the topics thoroughly.
 - E. Covered most of the topics thoroughly
29. Suppose you are solely responsible for the choosing the content in a regular high school mathematics course such as prealgebra, algebra, geometry, or algebra II. What would be your intentions concerning coverage of the statistical topics that appear in the textbooks for those courses:
- A. Completely skip all statistical topics.
 - B. Introduce a few of the topics.
 - C. Introduce many of the topics.
 - D. Cover some of the topics thoroughly.
 - E. Cover most of the topics thoroughly

VITA

Michael Scott Winstead was born on February 15, 1968, in East Liverpool, Ohio. He holds a Bachelor of Science Degree in Mathematics from East Tennessee State University, and a Master of Science Degree in Applied Mathematics from the University of Virginia. This dissertation completes the requirements for a Ph.D. in Education with an emphasis in Mathematics Education at the University of Tennessee, Knoxville.

Also certified to teach secondary mathematics, Mr. Winstead has taught in public high schools in Virginia and Tennessee. Since 1994, he has taught mathematics at Karns High School in Knoxville, Tennessee.