CONSTRUCTION OF AN INSTRUMENT DESIGNED TO ASSESS STUDENT ATTITUDES TOWARD THE POLICE

by

Paul Norlend Tinsley

B.A., Simon Fraser University, 1977

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS (EDUCATION)

in the Faculty

of

Education

SIMON FRASER UNIVERSITY

April 1988

APPROVAL

Name: PAUL N. TINSLEY

Degree: Master of Arts (Education)

Title of Thesis: Construction of an Instrument Designed to Assess Student Attitudes Towards the Police

Examining Committee:

Chairperson: Leone M. Prock, Associate Professor

M. Manley-Casimir
Senior Supervisor

D. Lingley
Visiting Professor

Carl J. Bognar
Bognar and Associates
204 - 3466 West Broadway
Vancouver, B.C. V6R 2B3
External Reader

Date approved April 19, 1988

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

CONSTRUCTION OF AN INSTRUMENT DESIGNED TO ASSESS STUDENT ATTITUDES TOWARD THE POLICE

Author:

(signature)

PAUL N. TINSLEY

(name)

18 Mar 1988

(date)

ABSTRACT

The attitudes of students toward the police, within the context of a democratic society, are worthy of careful consideration. However, a review of the literature reveals that the attitudes of elementary students are generally ignored, and that assessment techniques are generally inadequate, resulting in a great deal of contradiction and controversy. Because research is only as good as the methodologies and assessment techniques used, the purpose of this study was to construct an attitude scale (summated ratings type) to assess accurately the attitudes of students in grades 4 to 12 toward the police. Conceptually, this study assumes that the core of attitude is affect, and defines an attitude scale as a set of related opinion statements to which subjects respond. The psychological object of interest is defined as the attitudes of students toward the police, assessed by fixing attitudes along a quantitative continuum.

The sequence of scale construction is rigid, and the research design employed in this study incorporated both quantitative and qualitative approaches. First, a theoretical foundation on which to base opinion statements was defined, and a pool of statements was collected. Second, in order to establish readability, interviews were conducted with a small (12) non-random sample of students. Subsequently, based on the theoretical foundation and these interviews, a preliminary statement selection was made. Third, an item analysis, logical and empirical, was conducted, based on the responses of a

non-random sample of 245 students. Fourth, based on the responses of a new non-random sample of 246 students, reliability was estimated by means of the split-half technique and Cronbach's coefficient alpha. Finally, the issue of validity was addressed, but only to a limited extent. The result was an attitude scale that appropriately assesses the attitudes of students in grades 4 to 12 toward the police; however, further research needs to be done with respect to validity.

DEDICATION

To my wife

Pam

And my children

Corrina and Kelly

ACKNOWLEDGEMENTS

I wish to express my appreciation to the Faculty of Education, Simon Fraser University, for permitting me to pursue the research necessary to complete this study, with special recognition going to Dr. M. Manley-Casimir and Dr. D. Lingley for their assistance and support. In addition, I wish to express my appreciation to the staff and students of the school district of Abbotsford, British Columbia, for their cooperation in conducting this research.

TABLE OF CONTENTS

TABLES

CHAPTER I

Problem and Setting

Problem

The criminal justice system of most political systems includes three separate components: law enforcement, the judiciary, and corrections. This study is concerned only with the institution of law enforcement in a democratic society, particularly, England, the United States, and Canada. Law enforcement is the mandate of the police in any political system, but in a democracy it is unique in that the ability of the police to perform their duties, notwithstanding legislative authority, depends upon public approval, support, and willing cooperation. Considering this dependency, public attitudes toward the police may be a critical indicator of the state of law enforcement in our society, not only with respect to policing *per se*, but also with respect to the larger issue of social stability.

One of the primary concerns, then, of a democratic society should be the relationship between the police and the public, including its young people, because negative attitudes toward the police may begin at a very early age, justified or not, and persist throughout life. Such attitudes may be harmful, to the individual, to the community, and to society. In a democratic society, where policing is based on the principle "that the police are the public and that the public are the police," people need to know and understand the place of

law and law enforcement in the social structure of their community and society, and to know and understand their rights and responsibilities as citizens. Considering the importance of policing in a democratic society, the development of the attitudes of young people toward the police should not be left to chance. At the very least, society should afford young people the opportunity to develop attitudes that are informed by an accurate knowledge of the role of policing in their society.

Because most young people are found in school, where behavior patterns are molded and attitudes formed, and since schools are public institutions that are designed to serve society, it is incumbent upon schools to become actively involved in an attempt to educate youth regarding the role of the police in society. Included in the concept of education is attitude. According to Anderson (1981), affective outcomes are at least as important as cognitive outcomes in the educational process (p. 14); and Tyler (1973), quoted by Anderson, states that "human feelings are important both as a means and ends in education" (p. 2). The objection that schools (and police) should not be in the business of attitudes is at the very least naive. Philosophy, morals, and values are inherent in and inseparable from any curriculum--society decides what is of value and subsequently designs educational programs which promote these values, a process known as "political socialization."

The problem, though, is that attitude assessment is in somewhat of a morass. The problem starts with the conceptual

argument about what attitudes really are, and subsequently moves into the operational area of measurement. With respect to attitudes toward the police, problems in attitude assessment are further compounded by the fact that measurement instruments are often hastily constructed and are not grounded in well developed theory. As a result, research findings based on these instruments are suspect. In addition, elementary grade children are generally ignored in studies of attitudes toward the police.

Essentially, then, the heart of this study lies in the belief that the attitudes of young people toward the police within the context of a democratic society are of critical importance, and therefore deserve special attention, schools being the logical place for such attention. However, before the attitudes of students toward the police can be understood and informed decisions made, two issues must be addressed: (1) the attitudes of elementary students; and (2) inadequate assessment techniques.

Purpose

The general purpose of this study was to address the issue of public attitudes toward the police; the specific purpose of this study was to construct an attitude scale (summated ratings type) that would accurately assess the attitudes of students in grades 4 to 12 toward the police. The benefits of such a scale include facilitating research in the area of student perceptions and attitude change. Questions of interest here include the following: (1) why do attitudes deteriorate as grade level

advances? (2) do significant attitude changes occur at any particular grade? (3) what differences exist due to sex, socio-economic status, race, culture, religion? and (4) how stable are student's attitudes? Another benefit to be gained from such a scale is that it would facilitate the systematic evaluation of police-school liaison programs, a salient goal of which is to impact upon student attitudes. And of course, such a scale would facilitate research into what should be one of the primary concerns of a democratic society, the existing relationship between the public and the police.

Methodology

To address the problem of this study, the first task was to define the psychological object of interest. The whole area of attitudes is complex, marked by disagreement and controversy. This study does not attempt to resolve the controversy, but recognizes the importance of the relationship between the conceptual and operational definitions of attitude. Conceptually, this study defines the core of attitude as affect; and, operationally, defines an attitude scale as a set of related statements of opinions to which subjects respond. The attitude scale is entitled, "How Students Feel About the Police," and focuses on the attitudes of students in grades 4 to 12 toward the police, which is argued to be a unidimensional variable.

The psychological object of interest thus defined, the second task involved gaining access to appropriate samples. Because of practical limitations, the principles of random sampling were not used for this study. Rather, subjects (students) and subject groups (classes) were selected as access to schools was gained and cooperation with administration and teaching staff was obtained. Therefore, the samples were selected from the school district of Abbotsford, British Columbia, where the school board has approved a police-school liaison program for use in the schools. Because this study was concerned with student attitudes toward the police, research which is logically related and potentially beneficial to police-school liaison curricula, cooperation was relatively easy to obtain. However, limitations inhere in this study because the research depended upon the cooperation of the school district. For example, because the samples were not random, and because of the time factor, any argument for generalizability beyond the classes and schools involved and the times described is tenuous, and all conclusions presented in this study should be viewed within such a perspective.

The actual construction of the scale, which included both qualitative and quantitative methodological approaches, involved a number of designs within the overall design. Initially, the theoretical foundation of the scale was defined, and a large number of opinion statements that appeared to reflect the psychological object of interest were assembled. From this item pool, based on the defined theoretical foundation and scale

theory for summated ratings scales, a preliminary statement selection was made.

Prior to item analysis, because the scale is intended for students in grades as low as four, the issues of readability and ability of elementary students to differentiate between the responses as required on the standard summated ratings scale had to be addressed. Theoretically, readability level was established at grade four, as was the ability of elementary students to differentiate between the responses of a standard summated ratings scale. In order to operationally assess these theoretical conclusions, arrangements were made to conduct interviews with a small sample of students. Consequently, it was necessary to construct an interview guide, designed specifically for this study.

In order to conduct an item analysis, the next stage in scale development, a logical analysis of the statements was made and a prototype of the proposed scale was constructed. Subsequently, the prototype was administered to a second sample of students in order to facilitate an empirical analysis; i.e., a statistical analysis of the responses made to the opinion statements. Finally, based on these analyses and previous research, a final statement selection was made, which in effect yielded the finished version of the proposed scale.

To facilitate a preliminary evaluation of the new scale, it was administered to a third sample of students in order to obtain an estimate of reliability (through statistical measures, such as the split-half technique and Cronbach's coefficient

alpha). After an assessment of reliability, the evaluation of the new scale logically led into the issue of validity. Because of this study's limitations, scale validity, apart from a cursory assessment of empirical validity, was generally confined to an assessment of logical validity; i.e., face validity and content validity. In the case of attitude scales, validity is very difficult to establish, and in the developmental stage most researchers are satisfied with acceptable levels of reliability, leaving a demonstration of validity to subsequent research.

Organization

This chapter describes the problem, the purpose, and the methodology for this study.

Chapter II reviews the literature concerning attitude assessment, and attitude assessment with respect to the police.

Chapter III outlines the methodology used for this study, describing in detail the sample, the design and procedure, and the limitations of the study.

Chapter IV looks at preliminary instrument construction and initial statement selection. An important consideration at this stage is the instrument's theoretical foundation, which is described in some detail.

Chapter V examines the issues of readability and ability of elementary students to differentiate between the responses as required on the standard summated ratings scale, and concludes with a selection of statements appropriate for item analysis.

Chapter VI describes the process of item analysis as used in this study, and concludes with the statements that were selected for the finished scale.

Chapter VII outlines the procedures used to estimate the reliability of the final form of the attitude scale; validity is also briefly discussed.

Chapter VIII summarizes the procedures used to construct the proposed attitude scale, examines the final form of the scale with respect to the theoretical foundation, draws conclusions, and offers suggestions for future research.

CHAPTER II

A Review of the Literature

This review, for the purposes of comprehensiveness, includes not only the research regarding the attitudes of students toward the police, but also the research regarding attitudes in general toward the police and the **research regarding attitude assessment itself.** The structure of this literature review includes two major divisions: the first deals exclusively with attitude assessment; and the second deals with attitude assessment with respect to the police.

Attitude Assessment: A Look at the Issues

Dawes, in Fundamentals of attitude measurement (1972), said, "There is really no necessity that social psychologists agree about the definition of attitude in order to measure attitudes" (p. 16). This seems to be somewhat of a controversial statement, apparently questioning the concept of validity; however, this statement is not as radical as it might first appear, but is more a practical response to the confusion that exists in attitude theory. For example, conceptual definitions are numerous, complex, and often contradictory, a situation which may be described as intolerable (Scott, 1968, p. 265), particularly for anyone interested in attitude assessment.

In contrast to this conceptual confusion, however, one

finds some agreement in operational theory. The problem here, though, a problem inherent in Dawes' (1972) statement, is that there should be a strong relationship between the researcher's conception of the attitude under study and the measurement method used to assess this attitude, such a relationship crucial for interpreting the findings of a study (Lemon, 1973, p. 29).

The purpose of this review, then, is not to detail techniques for attitude assessment, but to examine some of the issues in attitude assessment, special attention being given to the theoretical constructs of attitude, the measurement process, and the adequacy of quantitative assessment.

Conceptual Definitions. According to McGuire (1985), the whole area of attitude is complex, confused by conceptual and theoretical problems, which includes contradiction and controversy, and according to Adams (1982), there are more conceptual definitions of attitude than of any other concept in social psychology. For example, the literature indicates that certain theorists enjoy periods of popularity, decline, and subsequent renewal (rf. McGuire, p. 240). Currently enjoying a period of popularity is the theory that an attitude is a "unifying mediational construct," an intervening variable between antecedent conditions and subsequent responses (McGuire, p. 240). Ajzen and Fishbein (1980), however, see this theory simply as an explanation proposed by Doob (1947) in response to the criticism that attitudes fail to predict behavior.

According to Ajzen and Fishbein (1980), the

"multicomponent" theory is predominant in attitude theory, which defines attitudes as complex systems incorporating a person's beliefs about an object (cognitive component), a person's feelings toward an object (affective component), and a person's tendencies to act or behave with respect to the object (conative component) (p. 26; rf. also Borg & Gall, 1983, p. 341). This theory, though, according to McGuire (1985), is not inconsistent with the "unifying mediational construct" theory, and is often subsumed by it (p. 242).

The prevailing conceptual definition notwithstanding, important alternate theories exist. For example, one theory defines attitude as only one of a number of "mediating dispositional variables" (McGuire, 1985, p. 241), such as interests, opinions, values, and preferences (rf. Ajzen & Fishbein, 1980; Allport, 1935; Anderson, 1981; and Rokeach, 1973). Apparently, much of this complexity in attitude theory was precipitated by Allport, who argued that the "evaluative dimension alone could not capture the complexity of the attitude concept" (Allport, cited in Ajzen & Fishbein, p. 17), and it is probably unrealistic, according to Scott (1968), to expect a satisfactory resolution, although Ajzen and Fishbein report a trend to return to the unidimensional definition of attitude, best stated by Thurstone (1931) as "the affect for or against a psychological object" (cp. Ajzen & Fishbein, pp. 26-27).

Operational Definitions. Conceptual definitions provide explanations about attitudes; operational definitions provide

practical methods to obtain information about attitudes. Thurstone and Chave (1929) provide an excellent example of the distinction between the two. In their book, The measurement of attitude, attitude is defined as "the sum-total of a man's inclinations and feelings ... about any specific topic" (pp. 6-7). In order to operationalize this conceptual definition, they decided to measure opinion, which in turn was operationally defined as a verbal expression of attitude: "Opinion has interest only in so far as we interpret it as a symbol of attitude. It is therefore something about attitudes that we want to measure. We shall use opinions as the means for measuring attitudes" (p. 7).

There is, of course, a discrepancy between the index of measurement and what is; i.e., the discrepancy between the attitude scale and the "real" attitude. This discrepancy, however, is universal for all situations involving measurement (Lemon, 1973, pp. 36, 117; Thurstone & Chave, 1929, p. 8), and it is in the degree of discrepancy that the importance of the relationship between the conceptual and operational definitions is found. According to Lemon, "The relation between the investigator's conception of the attitude he is studying and the method of measurement he uses to measure this attitude is of crucial importance in interpreting the findings of his study..." (p. 29; rf. also Anderson, 1981). The agreement between the measurement and what is will never be perfect, but the degree of agreement is important to the extent that the measurement will be considered valid; this is fundamental to measurement, whether

measuring temperature, gas, attitude, or anything else.

With respect to attitude, though, the problem is that there are more conceptual definitions of attitude than of any other concept in social psychology (Adams, 1982, p. 180). According to Adams, because of this conceptual confusion, a demonstration of validity in an assessment instrument is difficult, and it is no wonder, then, that the measurement of attitudes has developed independently of the theoretical controversy, and that the concomitant conceptual definition has tended to be quite narrow, a situation driven by a recognition amongst researchers of the importance of the relationship between the conceptual and operational definitions.

The Measurement Process. To assess usually means to determine quantitatively a specific dimension of something, which is generally quite objective when dealing with the concrete. However, when dealing with the abstract, objectivity is difficult to achieve. Most human characteristics, such as attitudes, are not directly observable, which is why an absolute definition of attitude is impossible; practically, we are limited to operational definitions. According to Adams (1982) and Anderson (1981), attitudes must be inferred from some kind of behavior; and at this point, it is important to note Cronbach's (1970) distinction between attitude and cognitive assessment: attitude assessment is concerned with "typical performance," behavior that is usually exhibited *vis-a-vis* maximum performance, which is the interest of cognitive tests.

Although many operational definitions of attitude exist, according to Lemon (1973), McGuire (1985), Scott, (1968), and Tittle and Hill (1967), there is a widely shared operational or working definition that is primarily concerned with affect. This definition describes attitudes as responses that locate objects of thought on dimensions of judgment (McGuire, p. 239). Objects of thought are foci of interest or objects of interest, which may be concrete or abstract; dimensions of judgment imply a mental continuum on which a person locates the object of interest.

According to Adams (1982) and Scott (1968), the procedures of attitude assessment can generally be classified into three stages: identification, collection, and quantification. Identification refers to describing the types of behavior responses that are acceptable for making inferences. Collection refers to the means by which the behavior responses are collected, usually a specially designed test or instrument. Quantification refers to the scoring of the obtained responses in order to collect meaningful information, the underlying assumption being that the score(s) will provide an index of the attitudes under study.

The fundamental issue here, of course, is using numbers or scores as indicators of a subject's responses, and ultimately his/her attitude. With respect to this issue, Eisner (1983) provides important criticism which helps to keep the limitations of quantitative assessment in perspective. For example, he states that "the single numerical test score is used to

symbolize a universe of particulars, in spite of the fact that the number symbol itself possesses no inherent quality that expresses the quality of the particular it is intended to represent" (p. 337; rf. also Eisner, 1985, pp. 12-24, for his criticism of quantitative testing). In addition, Anderson (1981) and Lemon (1973), although not proponents, raise the most devastating criticism of all, to which Eisner alludes, which is that any information that does not conveniently fit into the index imposed by the numbers is lost, which may be considerable, and as a result strict measurement results in such distortion that assumptions based on them are invalid.

Without accepting such an extreme position, which would necessarily require rejection of quantification for the purpose of attitude assessment, it must be recognized that quantification has a limiting effect on generalizations which may be generated by the measurement process. Here, Ajzen and Fishbein (1980), quoting H.C. Triandis (1967), make an astute point:

> There is a gap between those who are primarily concerned with the measurement of attitudes and those who have written theoretically about attitudes. The former frequently rest their case after providing us with a single score, whereas the latter make a large number of theoretical distinctions but do not provide us with precise and standard procedures for measurement. (p. 21)

Obviously, research is only as good as the methodologies and assessment techniques used (rf. Adams, 1982, p. 180; and

Shaw & Dobson, 1986, p. 382), but if the researcher is intent on waiting for a final definition of attitude and a resolution of all controversy, attitude assessment will never be possible.

Adequacy of Assessment. Within quantitative methodology, a number of techniques exist for the purpose of assessing attitudes, the popular ones being self-report techniques, observational techniques (which include physiological measures and unobtrusive measures), projective (partially structured stimuli) techniques, and multidimensional scaling techniques (Adams, 1982; Anderson, 1981). One reason for such a variety is differences between operational definitions; another reason is the search for a better technique--a closer agreement between the operational and conceptual definitions. For example, Ajzen and Fishbein (1980), arguing that because the core of attitude is affect, contend that the unidimensional, evaluative scales are appropriate to assess attitudes (pp. 54-55). Others, such as Lemon (1973), argue that such a narrow definition is indicative of measurement procedures defining attitude, rather than those procedures being devised to match the theoretical assumptions (p. 117).

There is no question, though, according to Lemon (1973) and Scott (1968), that most work in attitude assessment depends upon the self-report technique, which includes such methods a equal-appearing intervals (Thurstone & Chave, 1929), summated ratings (Likert, 1932), cumulative scaling (Guttman, 1944), and semantic differential (Osgood, 1952; Osgood, Suci, & Tannenbaum,

1957). According to Scott, such instruments are constructed to take advantage of the subject's self-awareness (assumption of ability), and the subject's readiness (assumption of willingness) to communicate verbally.

However, notwithstanding the dominance of the self-report technique in attitude assessment, general measures of adequacy are broad in their application; and according to Osgood (1952) and Scott (1968), any attitude assessment technique should include the following characteristics: (1) objectivity, in that any researcher should be able to apply the technique and obtain similar results; (2) validity, in that the technique measures what it purports to measure; (3) reliability, in that the results are repeatable; (4) sensitivity, in that distinctions can be made between persons along the quantitative continuum; and (5) utility, in that the instrument is relatively easy to administer, score, and interpret.

These characteristics, although their usefulness for determining the adequacy of attitude assessment techniques may appear self-evident to some, are quantitative in their orientation and are therefore based upon particular assumptions about attitudes and the measurement process (i.e., that attitudes <u>can</u> be quantitatively measured), assumptions that may not be shared by others, and are therefore subject to acceptance or rejection, depending upon the views of the researcher (rf. Scott, 1968, p. 251; and Smith, 1983, p. 9). For a detailed review of the controversy surrounding quantitative methodology and qualitative methodology and their respective philosophies,

the reader is referred to Smith's excellent article, "Quantitative versus qualitative research: An attempt to clarify the issue."

In conclusion, to say that there is no understanding of attitudes would be an overstatement, but the subject is so complex that to speak clearly of assessing an attitude is difficult. Rather, as Scott (1968) says, the researcher is restricted to a discussion on the "procedures for measuring a particular property of an attitude as conceptually defined" (p. 265). For this reason, most efforts in attitude assessment have concentrated, once the psychological object of interest has been defined, on measuring the properties of direction (positive or negative, favorable or unfavorable, etc.) and degree or intensity with respect to affect, the assumption being that the core of attitude is affect and that attitudes can be fixed along a quantitative continuum (e.g., self-report scales). The issues of conceptual confusion, agreement between conceptual and operational definitions, and the measurement process and its adequacy will probably never be resolved to everyone's satisfaction, which is the context in which the Dawes (1972) quotation must be understood; but for anyone interested in attitude assessment, these issues are important to consider.

The Assessment of Attitudes Toward the Police

Because of the important, sensitive relationship between the police and the public in a democracy, politicians, scholars, and the police themselves have serious concerns about police

image (Bouma & Williams, 1972, p. 119). Notwithstanding, one finds little research on the subject, either in the United States or Canada, which has been noted by American scholars such as Albrecht and Green (1977), Bouma and Williams, Klyman and Kruckenberg (1974), Portune, (1965), Torney (1971), and White and Menke (1978), and Canadian scholars such as Amoroso and Ware (1981), Griffiths, (1982), and Thornton (1975). The problem of limited research is further compounded by the fact that, as in the area of attitudes, the subject of attitudes toward the police is marked by contradiction and controversy. In addition, "the vast majority of the studies of public attitudes and perceptions of criminal justice concerns reflect no ... efforts to address" the problem of inadequate assessment techniques (White & Menke, p. 205).

The purpose of this review, then, is not to itemize studies on public attitudes toward the police, but to examine the existing state of affairs, special attention being given to contradiction and controversy, which pervades the research, the issue of the attitudes of youth toward the police, which plays an important part in the research, and the problem of inadequate assessment techniques, which plagues the research.

Contradiction and Controversy. To find contradictory research findings in the area of attitudes toward the police is not difficult. In order to put the issue in perspective, it is necessary first to outline the various independent variables of interest, given that attitudes toward the police is the

dependent variable, and second to examine the few areas in which general consensus exists.

First, the most common variables of interest are those of age (or grade, which according to Hess and Torney (1967) is highly correlated (r = .96) with age in the case of students), sex, race, socio-economic, ethnicity, geographic area, police contact (which includes police-school liaison programs), religious affiliation, minority groups, and intelligence. Other variables of interest include "dogmatism," situation in life or life satisfaction, environment, socialization process, "pluralistic ignorance," "ethics of social responsibility" and "ethics of personal conscience," and moral judgment. Attention, however, is not necessarily correlated with importance. For example, the work of Hess and Torney (1967) and Torney (1971) regarding the socialization process in young children is indeed a major study in the area of attitudes toward the police, even though this approach receives relatively little attention.

Second, general consensus exists with respect to the variables of age, race, ethnicity, and minority groups. For example, there is complete agreement that as age (or grade) in children advances, attitudes towards the police decline or become less favorable (cp. Amoroso & Ware, 1981; Bouma, 1969; Bouma & Williams, 1972; Hess & Torney, 1967; Lapsley, Harwell, Olson, Flannery, & Quintanna, 1984; and Portune, 1965). In addition, Derbyshire's (1968) definition of what constitutes a positive or negative attitude has never been questioned: people with positive or favorable attitudes toward the police perceive

them as being passive, kind, good, strong, non-authoritarian, hospitable, and friendly; conversely, people with negative or unfavorable attitudes toward the police perceive them as being aggressive, cruel, bad, weak, authoritarian, hostile, and angry.

Another example where agreement exists is with the variable race. Here, studies have shown that black people perceive the police more negatively than white people do (cp. Bouma, 1969; Bouma & Williams, 1972; Derbyshire, 1968; Klyman & Kruckenberg, 1974; and Portune, 1965). However, there is disagreement as to why black people and other minority groups often have relatively negative attitudes. Thornton (1975) suggests that the "immediate answer" is found in differential treatment by the police, but also suggests that negative attitudes may be due to groups who are not satisfied with their social position and see the police as an obstacle to their goals (rf. also Koenig, 1974). Portune (1965), however, states that there is no evidence to support the allegation that the negative attitudes of black youth are due to differential treatment by the police. According to Klyman and Kruckenberg, and Charles (1980), nonconformists and minorities often perceive the police to be oppressors, whose objective is to stifle dissent and harass, which accounts, at least in part, for the negative attitudes of these groups.

Regardless of the explanation to which one holds, the variable of police contact receives much of the attention in the literature, although it may be unjustified. Somewhat simplistically, some researchers tend to find the police

responsible for negative attitudes that may exist against them. According to Thornton (1975), "experiential factors" are the most significant in influencing attitudes toward the police, his thesis being "differential categorical treatment of the policed by the police" (p. 340; rf. also Rusinko, Johnson, & Hornung, 1978; and Chapman 1956, & 1960). Koenig (1975), however, states that "pluralistic ignorance," which occurs when people have unjustified beliefs based on an ignorance "of the behavioral, as opposed to verbal, activities of a large number of their peers" (p. 321; rf. also Bouma, 1969), may account for misperceptions of the police. According to Koenig, when people base their responses on their own experiences, as opposed to their perceptions of the experiences of others, their attitudes are more favorable.

The theory that blames the police for negative attitudes that may exist against them, a theory that hangs on the single variable "police contact," is rather limited though, and its generalizability is questionable. Realistically, to identify a single variable that can account for attitudes is impossible. According to Charles (1980), the evidence suggests that environmental factors, which include moral values held by people (rf. Lapsley, et al., 1984), outside the control of the police affect public attitudes toward them. This argument is supported by researchers such as Albrecht and Green (1977), who state that the studies concerned with single variables are limited by "their failure to consider that public attitudes toward police do not exist in isolation, but are part of a broader complex of

attitudes toward the system of legal justice..." (p. 67; rf. also Griffiths, 1982; and Torney, 1971). Their results show that negative attitudes toward the police are closely related to other components of the legal/political system, and that these attitudes are interrelated with feelings of alienation and powerlessness.

In addition to these complexities, there is a continuing argument as to what the proper role of the police in society, particularly a democratic society, is. Not only do people in general disagree regarding the role of the police, but also scholars, politicians, and the police themselves disagree. In addition, according to Portune (1965), widespread ignorance regarding general, operational policing procedure exists in our society. For example, the expectation that the police should be passive (rf. Derbyshire, 1968) is probably unrealistic, given the nature of the job at the street level. As a result, conflicting demands are placed on the police. This is the milieu in which the police find themselves; therefore, as stated by Albrecht and Green (1977), the police should bear only part of the responsibility for perceptions about them.

The Attitudes of Youth Toward the Police. The subject of attitudes of youth (the term youth includes children and adolescents, and also includes students, as most youth are found in schools in our society) toward the police has a special place in the general subject of public attitudes toward the police, as it raises interesting questions about attitude development and

attitude change. The youth of our society also has a special place in the criminal justice system, because in Canada 50% to 70% of police contacts are with youths under the age of 18 years (Griffiths, 1982, p. 329); and in the United States approximately 66% of all people arrested for serious crimes are under the age of 21 years (Beckman, 1980, p. 299). Finally, there is the critical issue of the socialization of youth into the political structure of our society. According to Hess and Torney (1967), the socialization process respecting youth has serious implications for the stability of our political system; and Torney's (1971) studies verify the importance of socialization "as predictors of the future adult political attitude matrix" (p. 153). As in all studies of attitudes toward the police, however, the studies of youth are conflicting (Radelet, 1973).

Phillips and Coates (1971) have succinctly stated the essential issue, "Among the nation's primary concerns are the existing relationships between the police and the citizenry" (p. 3). And no less is the concern with the existing relationship between the police and youth. For some time, there has been a general concern with what is perceived to be a negative and declining attitude by youth toward the police (rf. Bouma, 1969; Bouma & Williams, 1972; Portune, 1965, & 1968; and Radelet, 1973). However, other researchers suggest that no problem exists. For example, Griffiths (1982) states that "considerable evidence" exists to suggest that youths have favorable attitudes toward the police (p. 335; rf. also Amoroso & Ware, 1981).

Rather, he continues, it is specific groups that have negative attitudes. This is consistent with Derbyshire's (1968) research, who found problems with negative attitudes were confined to lower class Negro and Mexican youth.

The concern with negative attitudes of youth may be due in part to the steady decline of attitudes that occurs with advancing age (or grade, in the case of students). However, the evidence suggests that after a certain age is attained, such as that reached at the college level, the trend slowly begins to reverse (Lapsley, et al., 1984), and attitudes start becoming more favorable as the aging process continues (Koenig, 1974). This concern, then, may be misplaced, and should be redirected to how attitudes are formed, to how attitudes develop, and to how attitudes can can be influenced or changed.

Children are in a crucial stage of attitude development (Bouma, 1969; Derbyshire, 1968; Radelet, 1973; and Rusinko, Johnson, & Hornung, 1978), although some disagreement exists about what age is the most crucial and susceptible to influence. Griffiths (1982) asserts that the contacts youth have with the police are important "in shaping future relationships between adolescents and the criminal justice system" (p. 329), and it is this contact variable which has attracted a great deal of attention in attempts to influence the attitudes of youth. Pioneering research on this subject was conducted by Chapman (1956, & 1960), but the research conducted by Derbyshire was especially significant. Derbyshire found, surprisingly, that negative attitudes toward the police were formed very early

(grade one), but that they were often based on misperceptions and could be reversed through positive police contact, as found in police-school liaison programs.

Major research studies into the success of police-school liaison programs, a primary objective of which is to positively impact on student attitudes through positive contact with the police, have demonstrated that such programs are indeed effective (Bouma, 1969; Bouma & Williams, 1972; Portune, 1965, & 1968). Furthermore, according to Torney (1971), the police, as representatives of the legal system, have a major impact upon children. In Canada, apart from one study conducted in Ontario on students in grades 6 to 12 (Amoroso & Ware, 1981), which did not deal with the effect of police-school liaison programs, such research is nonexistent. However, unpublished research (rf. Ellis, 1973; Meehan et al, 1977; Pegler & Wright, n.d.; and Southwick, 1974) generally agree with the American research, concluding that police contact combined with educational programs do positively influence the attitudes of youth.

Notwithstanding these studies and their contribution to the research, the attitudes of youth progress through natural stages of development, previously alluded to. This process, known as political socialization, and although not understood very well, itself accounts for attitudes, positive and negative, toward the police. The major study here is by Hess and Torney (1967), who attempted to discover how the process of socialization takes place in youth. According to Torney (1971), because of the salient structural, formal aspects of the

home and the school in their apparent influence upon the attitudes of youth, researchers "easily fall into the error of a simplified, stereotyped picture of the process: The parents state the rules, the teacher provides the factual information about laws, both punish deviance, and the child absorbs and accumulates attitudes and modes of behavior" (p. 138); a tendency exemplified in police-school liaison programs.

Hess and Torney (1967) and Torney (1971) hypothesize that the complex interactions of four socialization models (accumulation, identification, role transfer, and cognitive development) may account for attitude acquisition. The cognitive development model, major contributors being Piaget and Kohlberg, is defined by Torney as a "sequential change in the organization of knowledge and the basis of judgment" (p. 140). For example, the older youth (e.g., teenager) has an increased ability to deal with abstract concepts and to reason, which has considerable influence on attitudes.

Other studies support this conceptualization of attitude development and acquisition. For example, Lapsley, et al. (1984), who studied attitudes toward authority in early and late adolescence, found important attitude development in adolescence, which they attributed to advances in "principled reasoning," which in turn explained the decline they found in the attitudes of youth toward authority. Young children see the power to punish and the power to control in a positive way; power is admired and associated with goodness (Torney, 1971). However, as children get older, they are able to separate power

attributes from other facets of a police officer's duties, and to think abstractly about the negative ramifications of power.

In conclusion, it appears that the deeper one looks into the issue of attitudes toward the police, the more complex the issue becomes. Notwithstanding the controversy and contradiction that has been examined, one issue yet remains, and this is the problem that plagues this particular research--inadequate assessment techniques.

Assessing Attitudes Toward the Police. As previously outlined, a great deal of contradiction and controversy exists in the research; but, regrettably, much of it is due to inadequate assessment techniques. In the desire to generate data on the subject, the means by which the data are generated has been relegated to a subordinate position, regardless of the fact that unless the generated data are based on a valid assessment, they are useless. However, this state of affairs does not seem to rate much consideration in the literature. The definitive study addressing this situation is by White and Menke (1978); but Charles (1980) has also made an important contribution. In addition, in a broader perspective, Gardner's (1975) critique of the recent research in attitude assessment is significant.

White and Menke (1978) point out the expanding effort to assess public perceptions of the police through attitude surveys (or questionnaires, or scales, these terms used synonymously), but that such scales are weak in a number of important areas.

First, through a critical examination of the assessment techniques used, one finds a misplaced faith in the assumption that attitudes can easily be measured. Such misplaced faith is demonstrated by the number of researchers who, in order to generate data to test their particular hypotheses, develop their own attitude scales, but report little evidence for such critical factors as item selection, reliability, and validity. For example, Klyman and Kruckenberg (1974) developed a "Citizen's Perception of the Police" scale for their study, but they give no descriptive information about their scale, except to say that the corrected reliability coefficient is .733. Similar deficiencies are found in a majority of studies that have attitudes toward police as their psychological object of interest (rf. Albrecht & Green, 1977; Amoroso & Ware, 1981; Bouma, 1969; Bouma & Williams, 1972; Koenig, 1974, & 1975; Marenin, 1983; Rusinko, Johnson, & Hornung, 1978; and Thornton, 1975), including unpublished studies (rf. Ellis, 1973; Meehan et al, 1977; Pegler & Wright, n.d.; and Southwick, 1974). Because evidence is not reported, it does not logically follow that a scale is invalid or inappropriate; but it is unacceptable not to report a scale's limitations, or to leave the reader to assume a scale is valid.

The assumption that attitudes can easily be measured is both naive and dangerous (White & Menke, 1978). As previously discussed, there is very little agreement on what attitudes are; and although the situation is somewhat better in the field of measurement, there is the problem of attempting to harmonize the

conceptual and operational definitions to support an argument for instrument validity. In addition, there are many problems inherent in the measurement process itself, such as the social desirability and the acquiescence response sets. Although wide consensus exists that such problems can be satisfactorily ameliorated, the "vast majority" of scales designed to assess attitudes toward the police, according to White and Menke, do not meet the necessary standards (p. 205).

In addition to failing to attend adequately to scale theory, which includes factors such as item selection, reliability, and validity, most researchers completely ignore the formulation and use of well developed foundational theory in scale construction. Well developed foundational theory is a key factor in scale development (Charles, 1980; White & Menke, 1978); and according to Gardner (1975), "effective research in this field [attitude measurement] demands the construction of attitude scales which clearly reflect some underlying theoretical construct..." (p. 101). After scrutinizing the literature in attitude assessment, he concludes that the major problems are lack of foundational theory, which he calls the "underpinnings" of the scale, and lack of appreciation of what the issues in attitude assessment are. These criticisms equally apply to the research in criminal justice; and not only to the research found in journals and texts, but also to major U.S. government studies, such as the President's Commission on Law Enforcement and Administration of Justice in 1967 (White & Menke).

Considering these issues, to find contradiction and controversy in the literature is inevitable. If, in their desire to generate data and report significant results, researchers depend on scales for which minimal evidence of the critical factors necessary for valid scale construction exists, the data are bound to vary and to conflict. The fact that the studies vary and conflict may in itself be prima facie proof that the scales used are inadequate, as they lack convergent validity (White & Menke, 1978). Although some variance in results may be explained by population differences, the extent of the conflict indicates problems at a basic level.

Notwithstanding, researchers generate data by means of suspect scales, apply statistical procedures, and offer interpretations, which other researchers "take at face value," even in literature reviews (White & Menke, 1978, p. 215). According to Gardner (1975), this is absurd, stating that the absence of well developed foundational theory makes subsequent statistical procedures "completely irrelevant" (p. 102). Currently, an unjustified emphasis is placed on statistics to the exclusion of theory and logical analysis (rf. Eisner, 1985). Regardless of the fact that complex statistical procedures are applied to data and that significant results are obtained, the findings are meaningless if the data are based on suspect scales--a house built of the best materials by the most skilled professionals using the finest techniques, on a foundation of sand. As a result, according to White and Menke, research findings in the field of criminal justice are at the very least

suspect, and "must be considered with caution" (pp. 205-206).

In conclusion, it has not been the intent of this review to attack particular researchers and studies, because no useful purpose would be served. The purpose of this specific part of the review, however, was to raise the issue of adequacy in attitude assessment, an issue for which few others have shown concern. According to Charles (1980), in order to ameliorate the problem, researchers who develop their own scales should carefully attend to theory and to specifics in design, clearly define the psychological object of interest, report limitations, and continually update; the alternative is the perpetuation of the problem and the corollary that research in the field of criminal justice will not be taken seriously .

Summary and Conclusions

The purpose of this chapter was to review the literature specific to attitude assessment, assessment of attitudes toward the police in general, and assessment of the attitudes of students toward the police. With respect to attitude assessment per se, an attempt was made not to fall into the trap of simply itemizing techniques. Here, as Gardner (1975) implies, what is important is to become aware of the issues in attitude assessment, and its inherent problems and limitations. With this purpose in mind, issues such as the contradiction and confusion found in the attempt by theorists to conceptually define attitudes and the problems that this creates for those involved in measurement were examined. The point is that the

relationship between conceptual and operational definitions is of no small significance--if no understanding of what is being assessed exists, assessment becomes trivial at best, absurd at worst. This of course, raises the question of validity, which can be extended into the question of reducing a person's attitude into a numerical score. These issues will probably never be resolved, but are critical in attitude assessment.

With respect to the general assessment of attitudes toward the police and the specific assessment of the attitudes of youth toward the police, an attempt was again made not to itemize the studies. Rather, an attempt was made to examine the existing state of affairs in the literature, special attention being given to contradiction and controversy which pervades the research. After examining the literature, it was found the much of the contradiction and controversy is probably due to inadequate assessment techniques, which seems to plague the research. The problem is that if the data are generated from suspect scales, subsequent statistical analysis is irrelevant and the findings are useless. What is needed, then, if research in the field of criminal justice is to be taken seriously, is for researchers to attend first to the basic issue of assessment techniques, which is the purpose of this study.

CHAPTER III

Methodology

Recognizing the need for valid measurement instruments in the field of criminal justice, the purpose of this study was to construct a summated ratings scale that would accurately assess the attitudes of students toward the police. This chapter provides an overview of the methodology used to accomplish such a purpose. First, the sample is described; second, the design and procedure of the study is outlined; and third, the limitations of the study are defined.

Sample

Description of the Area. The principles of random sampling were not used for this study because of practical limitations. Rather, subjects and subject groups were selected as access to schools was gained and cooperation with administration personnel and teaching staff was obtained. The schools, located in urban/residential areas, were selected from the school district of Abbotsford. One important, practical reason for selecting this district was that the school board had approved a police-school liaison curriculum; therefore, cooperation was relatively easy to obtain because the proposed research was logically related and potentially beneficial to this type of curriculum.

The school district of Abbotsford includes the municipalities of Matsqui and Abbotsford and is located in the

central Fraser Valley. Although Abbotsford and Matsqui have separate municipal governments, they form a "single socio/economic unit" (Matsqui, 1987, p. 5), and for the purposes of this study, it is useful to consider these municipalities as such. For example, these municipalities are described as politically "conservative," a stronghold of the provincial Social Credit and federal Conservative parties, and as somewhat religious, described in the media as the "Bible belt" of the Fraser Valley (The News, 1987).

In 1986 the population of Matsqui was 51,065 and the population of Abbotsford was 14,368; and it is estimated that in 1991 the population of Matsqui will be 62,130 and that the population of Abbotsford will be 17,480 (Matsqui, 1987, p. 10). Because Matsqui is by far the larger municipality with almost 80% of the population of the school district, the following descriptions will be confined to Matsqui; however, because these municipalities form a "single socio/economic unit," it can generally be assumed that any general description of Matsqui also applies to Abbotsford.

The main business area of Matsqui is called Clearbrook, which comprises approximately two-thirds of the urban concentration. Clearbrook is the principal town centre for the central Fraser Valley, with a trading population in 1975 of 113,000, estimated to reach 175,000 by 1991 (Matsqui, 1987, p. 38). Much of the growth of Matsqui, evidenced in population and trading population, can be attributed to its proximity to the city of Vancouver. This, combined with favorable housing

prices, rural environment, close proximity to the American border, has resulted in a large residential population who live in Matsqui but work outside the district.

The average household size in 1981 in Matsqui was 3.1, and 33% of the population was under 19 years of age (Matsqui, 1987, pp. 12-13). Residential housing is predominantly single family, but recently a shift to "more of a diverse mix of high and low density housing types" has occurred (Matsqui, p. 20). For example, in 1971 multi-family units accounted for only 6% of housing, but in 1986 multi-family units accounted for 30% of housing (Matsqui, p. 20). In 1980, the average family income in Matsqui was $25,975.00, compared to $28,798.00 for Abbotsford, $26,963.00 for Langley, $25,063.00 for Mission, and $31,871.00 for greater Vancouver (Matsqui, p. 16).

Agriculture is of major importance to both municipalities, described as one of the "most productive agricultural areas in Canada" (Matsqui, 1987, p. 1). Most prominent in agricultural activities are dairy, beef, egg and poultry, berry, and vegetable farming. Industry related to residential construction is also of major importance, Matsqui being a major supplier of sand and gravel in the Lower Mainland.

The school district of Abbotsford includes 31 elementary schools, located in both rural and urban areas, 5 secondary schools, located in urban areas only, 2 special schools, and a number of private schools, most of which are found in Matsqui. The schools selected for this study were all found in the urban/residential area of Matsqui, which means, with respect to

the sample, that the elementary group was drawn from the urban area only, and that the secondary group was drawn from both the urban and rural areas.

Description of the Subjects. The methodology of this study necessitated gaining access to subjects/groups on three separate occasions. First, in order to operationally assess readability of the scale which had been theoretically established, it was necessary to conduct interviews with a small sample of students who were selected from two schools, an elementary school and a junior high school. A necessary criterion, however, was that the interview be entirely voluntary, students not coerced to participate in any way; therefore, the students of necessity were willing and cooperative. A brief profile was conducted on each student, considering characteristics such as intelligence, academic background, and representativeness of peer group (rf. Appendix A).

Specifically, then, four students, two males and two females, all Caucasian, were selected from grades four, six, and eight. These students were generally described as "average," with some exceptions, however, which might be found in any "average" classroom. For example, in grade four a student with low reading ability was included, along with a student described as "high average"; in grade six a student who had difficulty with new material, a shy student, and a student described as a "bit above average" were included; and in grade eight an

"average" student who worked "really hard" and one who was a "lower achiever" were included.

Second, with respect to conducting an item analysis and estimating reliability, which required gaining access to a sample of students on two separate occasions, a profile was conducted for each class (rf. Appendix B and D), considering characteristics such as race (approximately 10% of classes in the school district of Abbotsford consist of East Indian students), intelligence, academic performance, behavior, attitude, geographic area, family background, and any unusual characteristics that would distinguish the class from other classes in the school district. In addition, the prototype used for item analysis (rf. Appendix C) and the final version of the scale (rf. Appendix E) included a provision to record the sex of the respondents, which was generally evenly distributed. These classes, then, selected from two elementary schools, two junior secondary schools, and two senior secondary schools, were described by the teachers and principals involved as "average" within the school itself and also the school district.

Description of Sample Size. With respect to the interviews, a small sample of 12 students was selected, sample size being one of the limitations of interviews because of the time required to conduct an interview. Although it is difficult to generalize from such a small sample, one of the main objectives of the interviews was to prove or disprove theoretical conclusions already reached, an objective which was

met. In addition, empirical analyses (i.e., item analysis and reliability tests) verified conclusions reached as a result of the interviews.

With respect to the item analysis and reliability studies, one class each of grades 4 to 12 for each study was selected, the purpose being to administer the prototype and the finished scale to a sample of the target population. To have conducted an item analysis and estimated reliability for each grade would have been ideal; however, practical limitations prevented this. For example, to conduct such studies the sample must be sufficiently large for statistical purposes, the recommended, minimum sample size for item analysis being five times the number of statements constituting the prototype (Anderson, 1981, p. 248; Scott, 1968, p. 259). Therefore, for the purposes of this study, where the prototype consisted of 17 statements, a sample size of at least 85 (17 x 5) was necessary, although Borg and Gall (1983) recommend a sample of at least 100 subjects (p. 299), and Adams (1982) recommends a sample of 100-200 subjects (p. 181).

Sample sizes are not specifically recommended for estimates of reliability, but a review of the literature indicates that researchers, developing summated ratings scales, use a sample that is approximately the same size as that used in item analysis. For example, Edwards (1957) used 72 subjects for item analysis and 80 subjects for estimating reliability; and in other examples, where summated ratings scales were constructed for the purpose of assessing attitudes toward the police,

reliability was estimated from sample sizes ranging from 72 (Chapman, 1960) to 100 (Portune, 1965) to 117 (Phillips & Coates, 1971). According to Borg & Gall (1983), "the general rule is to get the largest sample possible" for any statistical analysis, because the larger the sample, the more likely the results will be representative of the general population (p. 257).

Although this study did not analyze each grade level, the sample was representative of the target population (students in grades 4 to 12), and was sufficiently large for item analysis (N = 245) and for estimating reliability (N = 246). In addition, because of the relatively large size of the total samples, they were divided into elementary groups (N = 111 for item analysis and N = 116 for estimates of reliability) and secondary groups (N = 134 for item analysis and N = 130 for estimates of reliability) in order to ascertain whether or not the results of the analysis of the total group were consistent with what was happening in the elementary and secondary groups. If the proposed scale is not appropriate for the range of grades for which it is intended, this should be reflected in the empirical data (Anderson, 1981, p. 122).

Traditionally, the school system has separated elementary and secondary students because of the significant changes that occur in a person who is approximately 12 or 13 years of age. For example, Jean Piaget identified children in the 6/7 to 11/12 age range as being in the stage of "concrete operations" and children in the 11/12 and on age range as being in the stage of

"formal operations" (rf. Weiner & Elkind, 1972, p. 280). The age at which children reach these stages are dependent of course upon maturation factors, but are also dependent upon background and experience (Weiner & Elkind, p. 280; rf. also Parker, 1984, pp. 18-28).

The result, therefore, of first examining the responses of the total sample, and of secondly examining the responses of the constituent elementary and secondary groups, was increased confidence in deciding upon the best statements for inclusion in the finished scale, in estimating reliability, and in concluding upon the appropriateness of the proposed scale for the intended grades.

Design and Procedure

Construction of a psychological measure is very ordered, and can be explicated in a description of the necessary steps (rf. Borg & Gall, 1983, pp. 298-301). First, it is necessary to define the objectives; i.e., to describe the population of interest, the trait to be measured, the technique selected for the measurement process, and the scoring process. Second, a review of the relevant literature is helpful in order to isolate the issues and to provide valuable information about similar measures. Third, the theory on which the scale is based should be clearly outlined, followed by a description of the development of an item pool. Fourth, after a suitable sample of items has been selected, which in itself may involve a number of steps, such as establishing readability level, a prototype of

the instrument should be prepared and administered to a sample representative of the population of interest. Item analysis is conducted at this stage in order to facilitate a final selection of statements; in addition, further refinements may be made to the format of the scale at this stage. And fifth, on the basis of the previous steps, the final version of the scale is prepared, and again administered to a new sample from the target population. At this point, data is collected in order to estimate reliability and assess validity, permitting conclusions to be made about the scale.

Trait To Be Measured. According to Dawes and Smith (1985), during the seventies more than 20,000 articles and books were listed under the general heading of attitude in a single psychological journal, the aggregate marked by conceptual and theoretical problems. This study has no intention of attempting to define attitude in any absolute way; however, the importance of the relationship between the conceptual and operational definitions of attitude is recognized.

Conceptually, this study holds to the argument that the core of attitude is affect, a unidimensional variable that is synonymous with feeling in an emotional sense (rf. Ajzen & Fishbein, 1980; and Edwards, 1957, p. 2). Because attitudes are not directly observable, which is why an absolute definition is rather an elusive endeavor, attitude assessment is limited to an operational definition, such a definition dependent upon the assumption that attitudes can be inferred from certain kinds of

behavior (Adams, 1982, p. 180; Anderson, 1981, p. 28); i.e., in the case of attitude scales, written responses to statements of opinion (Dawes & Smith, 1985, p. 510; Lemon, 1973, p. 178; Likert, 1932, p. 9; Murphy & Likert, 1967, p. 3; Thurstone & Chave, 1929, p. 7). For the purposes of this study, the proposed attitude scale is defined as a set of related statements of opinions to which subjects respond, providing meaningful information about the psychological object of interest (rf. Anderson, p. 78).

The proposed attitude scale is entitled "How Students Feel about the Police" and has as its interest the attitudes of young people, particularly students in grades 4 to 12, toward the police. The study proposes that a general attitude toward the police exists, even in grade levels as low as four, and that it is a unidimensional variable (rf. Bouma & Williams, 1972; Chapman, 1960; Derbyshire, 1968; Goran, 1970; Phillips & Coates, 1971; and Portune, 1965). Therefore, a unidimensional, evaluative scale, such as the summated ratings scale, is appropriate to assess student attitudes toward the police (rf. Ajzen & Fishbein, 1980, pp. 54-55).

Technique Selected. Given that a unidimensional, evaluative scale is appropriate to assess students attitudes toward the police, the particular technique selected was the method of summated ratings, developed by Rensis Likert (1932). In addition to its obvious appropriateness, this technique was selected for a number of reasons. First, it is relatively

simple to construct, as compared, for example, to the equal-appearing intervals method (rf. Thurstone & Chave, 1929), and the cumulative method (rf. Guttman, 1944) (Anderson, 1981, p. 87; Borg & Gall, 1983, p. 342; Edwards & Kenney, 1967, p. 255; Likert, p. 42). Second, it is "relatively simple for subjects to understand and use" (Phillips & Coates, 1971, p. 4). Third, it is generally as reliable as other attitude scales (Adams, 1982, p. 181; Borg & Gall, p. 342; Edwards, 1957, p. 162; Edwards & Kenney, p. 255; Likert, p. 42; Phillips & Coates, p. 4; Seiler & Hough, 1970; Tittle & Hill, 1967). Finally, in a study of constructing scales with respect to attitudes toward the police, Phillips and Coates found that respondents "preferred" the Likert-type scale over the semantic differential (p. 12).

The psychological object of interest defined and the technique selected, the proposed scale is concerned with the properties of direction (i.e., favorable or unfavorable, positive or negative) and degree (i.e., intensity or magnitude), the assumption being, of course, that attitudes toward the police can be fixed along a quantitative continuum.

Scoring. Likert's technique is often called the "method of summated ratings," a term coined by C. Bird (1940) to describe Likert's scoring procedure (rf. Edwards & Kenney, 1967, p. 249). "Summative" designates a scale that is scored by adding the response scores of each item (Scott, 1968, p. 218). For example, values of one to five are assigned to each of the

five responses found in the standard Likert-type scale: one is assigned to the negative extreme of the attitude continuum, and five to the positive extreme of the attitude continuum (assigning one vis-a-vis five to the negative extreme is, of course, purely arbitrary, but is common practice). After the numerical values are assigned to the responses, the score for each subject is determined by adding the numerical values for each response (Likert, 1932, pp. 25-26; Edwards & Kenney, p. 250). The properties of direction and degree, then, with respect to the attitude of interest, are represented in the values assigned to each response for each statement, and are subsequently reflected in the total score.

A criticism of summative scoring is that the scores cannot be interpreted independently of the scores of some defined group (Edwards, 1957, pp. 156-157). However, this is only a problem if the individual is the focus of interest; if the purpose of the scores is to make comparisons between groups, which is the general purpose of the proposed scale, "Likert-type scaling offers no problems" (Edwards, p. 33). Notwithstanding the practical limitations which Edwards correctly identifies, Anastasi (1976) more accurately points out that if norms are established, the scores of Likert-type scales are also meaningful with respect to individuals.

Scoring, then, is simply the quantification of obtained responses in order to collect meaningful information, the underlying assumption being that the score(s) will provide an index of the attitude under study. As described, all items of

the scale are used to derive the total score, which assumes that the scale is unidimensional; i.e., that the scale does not have subscales, therefore permitting attitudes to be fixed along a single quantitative continuum. This assumption is important, for only then is one justified in combining the responses to all the statements into a single score. For example, if I wished to measure someone's height, I would not sum height and weight because the single variable height would be confounded by the variable weight. In other words, it is a mistake to identify all the attributes of something and to subsequently sum them altogether, thinking the score will be meaningful. The procedure of summated ratings, therefore, assumes unidimensionality and implies that each statement is monotonic, or a linear function of the same attribute (Dawes & Smith, 1985, p. 532; Scott, 1968, pp. 219, 247). Without claiming to go beyond what Scott describes as the "modest achievement" (i.e., for most subjects, the total score represents, for the most part, the intended trait) with which most researchers are satisfied (p. 250), this is the assumption made for the opinion statements found in the finished scale in this study.

One final point that is important to note with respect to scoring in summated ratings scales is that the numerical values assigned to the responses are not interval in nature, although Likert (1932) makes an argument for this (pp. 21-25, 42), but are ordinal with respect to the properties usually associated with numeric data (Dawes & Smith, 1985, p. 532). Strictly speaking, only "order-preserving transformations" are possible,

but all forms of statistical analysis are performed on a regular basis (Dawes & Smith, pp. 532-534; Hess & Torney, 1967, p. 239; Moore, 1979, pp. 121-123). The question then arises concerning the validity of such operations; but most quantitative researchers, such as Dawes and Smith, express only "slight concerns," and according to Erickson and Nosanchuk (1977), "pragmatic investigations have shown that ordinal data can be treated like interval data pretty safely if the data have a fairly smooth distribution, N is fairly large, and the test is robust" (p. 167).

This study attempted to address these "ifs": attitudes were assumed to be distributed normally in the population (rf. Likert, 1932, p. 22), the sample size met or exceeded the standards found in the literature, and the tests and procedures used were those recommended or used by prominent researchers in attitude assessment. Notwithstanding the fact that the recommended criteria were met, to be aware of the assumptions involved in applying the interval/ratio properties of numbers to essentially ordinal data is important.

Theory. Theory played an important dual role in the development of this scale. First, each opinion statement was chosen on the basis that it reflected adequately the domain of interest, which was based on well developed theory (i.e., foundational theory). For example, a major premise of this study is that policing in a democracy is dependent upon public approval, support, and willing cooperation, which has its roots

in the principles of Sir Robert Peel. This theory formed the conceptual underpinnings, the foundation, of the scale, and remained a critical factor in every stage of development. Second, scale theory itself was a critical factor in the development of this scale. For example, scale theory for summated ratings scales includes criteria for statement selection, item analysis, estimating reliability, and assessing validity. Foundational theory, then, preceded scale theory, but both were indispensable in the construction of the proposed scale.

Item Pool. A pool of statements, 79 in total, was obtained from a number of studies that included sentence based scales constructed to assess attitudes toward the police (rf. Hess & Torney, 1967, p. 233). From this pool, a small sample of 19 statements was selected on the basis that it met the criteria for inclusion in summated ratings scales, and that it appeared to reflect adequately the universe of interest in that it was consistent with the described theoretical foundation. Subsequently, this sample was refined as a result of rewriting the statements in order to establish readability.

Readability. Given the pencil and paper nature of self report instruments, and the fact that the Likert technique is based on a series of sentences, and given that this instrument is intended for students in grades as low as four, it was necessary to reduce the readability level to that of grade four,

which was accomplished on a theoretical level, subsequently tested on a practical level through personal interviews with students in grades four, six, and eight. As suggested by Anderson (1981), a "comparison of estimates" (obtained by using the formulae of Dale and Chall (1948; Instructions, 1948), Fry (1968; 1977), Rowls (rf. Anderson, p. 99), and the word lists of Dale (Dale & Chall, Instructions, pp. 45-54), and Allington et al (1985)) was used to establish readability at the theoretical level (p. 101). A related issue was whether or not students in grades as low as four have the ability to differentiate satisfactorily between the responses as required on a five-point summated ratings scale, which was also assessed through the personal interviews.

In order to conduct the interviews, arrangements, which included an informative letter for the parents of those students selected for the interviews, were made to conduct the necessary research in the school district of Abbotsford, British Columbia. The elementary students were selected by the principal and teachers involved, and the junior high students were selected by a counselor, the goal being to select students who were representative of their peer groups in an "average" sense. These interviews, conducted privately in order to avoid unnecessary distractions and interruptions that might influence the results (Borg and Gall, 1983, p. 439; Parker, 1984, p. 21), were standardized by means of an interview guide developed specifically for this study (rf. Appendix A).

An important point is that the interviews were not

designed to assess the attitudes of the students interviewed, but were designed to facilitate the construction of the proposed attitude scale, and proved to be a reading exercise for the students. Notwithstanding, the students' right to confidentiality was respected (Borg & Gall, p. 443; Parker 1984, pp. 24-26), of which they were apprised. In addition, the students were fully informed of the nature and purpose of the interviews (Farley, 1981, p. 185; Gross, 1984, pp. 67-68; Parker, p. 27). Finally, the dynamics of the interviews were tailored to the mental, emotional, and physical comfort of the students, ensuring that they were relaxed and that the atmosphere was nonthreatening (Baker, 1983, pp. 512-515; Borg & Gall, p. 440; Farley, p. 185; Parker, pp. 19, 21).

The Student Interview Guide (rf. Appendix A), a critical component of the interviews, found its direction in three objectives. The first objective of the interviews was to assess the readability level, theoretically established at grade four, of the statements selected from the item pool. In order to meet this objective, each student was asked (1) to read each statement aloud, (2) whether or not s/he understood the words in each statement, (3) the meaning of the statement, and (4) whether or not his/her peers would understand the statement. As a result, assessment for each statement was achieved by noting the number of problems students had with oral reading, vocabulary, and comprehension. In addition, by asking students whether or not their peers would experience problems with the statements, students were allowed to disassociate themselves

from the question and feel free to offer an objective opinion. By examining the Student Interview Guide, it is apparent that the format is standardized and semi-structured, the purpose being to facilitate objectivity and reliability (Borg & Gall, 1983, p. 442; Kerlinger, 1964, p. 470). In addition, in order to prevent the possibility of built-in bias, the order of the statements was determined by random assignment, using a table of random digits according to accepted procedures (rf. Moore, 1979).

Continuing with the first objective, with respect to the readability of the responses, each student was asked (1) to read the words aloud, (2) whether or not s/he understood the words, (3) whether or not his/her peers would understand the words, and (4) to define the word "unsure" (rf. Chapter V). As a result, assessment of the responses was achieved by noting the same factors noted in the assessment of the statements.

The second objective of the interviews was to determine whether or not students in grades as low as four have the ability to differentiate between the responses as required on a five-point summated ratings scale. Here, for comparison purposes, the participation of grade eight students was especially important because their ability to respond to the five-point Likert scale is generally unquestioned. In order to meet this objective, a decision was made to involve the students in a role playing game, a tactic suggested as appropriate in interviews with children (Farley, 1981, p. 186; Parker, 1984, p. 27), and appropriate for the preliminary evaluation of opinion

statements (Edwards, 1957, p. 13). Here, the students were asked to play five different roles, based on the five responses found in the standard summated ratings scale and on the domain of interest identified for this study. For example, the roles included the following fictional characters: one who really liked the police; one who liked the police; one who did not know how s/he felt about the police; one who did not like the police; and one who hated the police.

In role playing, the potential for interviewer bias was strong; therefore, the interview guide was designed so that, first, each student selected a statement of his/her own choosing; second, each student assumed a randomly assigned fictional role; third, each student read the selected statement quietly; and fourth, each student subsequently responded to the statement as if s/he were the fictional character. The beauty of such an exercise was that in addition to reducing the potential for interviewer bias, it prevented the students from perceiving, or falling into, a pattern and identifying the correct responses by simply reciting the sequential order of the responses found in the standard summated ratings scale. This process was repeated five times, so that each student had an opportunity to play all five roles. The assumption was that if the sample of students selected for the interviews could consistently respond in a manner consistent with a randomly assigned role, students in grades as low as four have the ability to differentiate between the responses of the standard summated ratings scale.

Following the interviews, the results were analyzed, and are reported in this study. The analysis is generally descriptive; nonetheless, it is informative for the purposes of this study, meeting the third and final objective of the interviews, which was to facilitate the revision and selection of statements that would be most appropriate for item analysis. Therefore, based on theory and the results of the interviews, the best statements were selected for the next step, item analysis.

Item Analysis. Inherent in attitude scales are the assumptions of willingness and ability on the part of the subjects to respond to the opinion statements, and within the assumption of willingness lies two problems which must be attended to in order to provide reasonable assurance that the subjects are responding honestly (Anderson, 1981, p. 65; Scott, 1968, p. 215). These two problems are often categorized in terms of response sets: the social desirability response set, and the acquiescence response set (Adams, 1982, p. 182).

The social desirability factor is the most common reason for a subject not responding honestly to the opinion statements. Here, people will tend to respond in a way that they believe is socially acceptable, rather than in a way that is consistent with their true beliefs and feelings (Anderson, 1981, pp. 65-67). Notwithstanding, research has shown that the social desirability response set can be minimized by the careful administration of the attitude scale. The following is a

synopsis of administration procedures recommended in the literature (rf. Adams, 1982, p. 182; Anderson, pp. 67-68; Borg & Gall 1983, pp. 304, 424; and Scott, 1968, pp. 215, 236).

1. The administrator should develop a rapport with the subjects.
2. Honesty should be actively encouraged.
3. There should be repeated assurance that the scale is not a test, and that there are no right or wrong answers.
4. Confidentiality must be assured (not necessarily anonymity).
5. The importance of the research should be emphasized.
6. Conditions should be conducive to honest responses (e.g., relaxed atmosphere).

The problem, then, is how to motivate the subjects to respond to the opinion statements in a typical, honest manner, a problem which can be adequately resolved through careful administration procedures. Obviously, a key player in this is the person who administrates the scale, and it is recommended to use persons whom the subjects trust and respect (Scott, 1968, p. 215), such as teachers in the case of students (Borg & Gall, 1983, p. 303).

These criteria, therefore, were carefully incorporated into a short survey manual (rf. Appendix B and D), the purpose of which was to attend to the problem of response sets and standardize administration. Another advantage to the manual is that it is a double-edged sword: the instructions, by

emphasizing the importance of subject cooperation, subtly influence the administrator in like manner.

The acquiescence factor refers to the tendency of the subjects to "react in terms of the format of the item, regardless of the specific content..." (Adams, 1982, p. 182). Acquiescence increases when the subject is unsure of his/her feelings, and when the scale is lengthy (Anderson, 1981, p. 67). In addition, children apparently exhibit a marked tendency to fall into this response set (Adams, p. 182). What appears to be the case, though, is that acquiescence is for the most part a function of instrument construction (Scott, 1968, p. 237), and can be satisfactorily resolved by simply using an equal number of positively worded and negatively worded opinion statements, and distributing them randomly throughout the scale (Adams, pp. 181-182; Anderson, p. 67; Lemon, 1973, p. 80; Likert, 1932, p. 46; Scott, p. 239). In addition, relatively short scales are recommended for children (Anderson, p. 69). Therefore, these recommended techniques were incorporated in the construction of the prototype (rf. Appendix C) and the finished scale (rf. Appendix E).

After a suitable sample of statements had been identified, and after the technical problems of readability and response sets had been satisfactorily addressed, a prototype of the instrument was prepared and administered to a sample of the target group for the purpose of conducting an item analysis (rf. Borg & Gall, 1983, p. 299; Chapman, 1960, p. 17; Lemon, 1973, p. 178; and Likert, 1932, p. 46). Item analysis, essentially a

formal process of examination where each opinion statement is analyzed regarding its suitability for inclusion in the final form of the attitude scale, was addressed in this study by means of a logical item analysis, and an empirical item analysis, using a simple correlational technique whereby the item score was correlated with the battery score, and using Likert's "criterion of internal consistency."

In order to test run the prototype, arrangements were made to conduct the necessary research in the school district of Abbotsford. The test run of the prototype (i.e, the administration of the prototype scale) was conducted in a regular classroom setting by volunteer teachers, who, in consultation with their school principals, were responsible for selecting the appropriate classes. Administration was standardized by means of the survey manual developed specifically for this study (rf. Appendix B). An important point is that the purpose of the prototype was not to assess the attitudes of the student respondents, but to facilitate the construction of the proposed attitude scale. Notwithstanding, the students' right to confidentiality was respected, anonymity being guaranteed, of which they were apprised. In addition, the students were fully informed of the nature and purpose of their participation.

Although item analysis was central, the test run of the prototype also addressed the issues of administration and communication value. Administration is a critical issue in attitude assessment because attitude scales are dependent upon

the assumptions that the subject has the ability to communicate his/her feelings with respect to the psychological object of interest and that the subject is willing to communicate these feelings verbally (Scott, 1968, p. 214). As previously discussed, proper administration can offer some assurance that the problem of the social desirability response set is resolved, and proper administration can be reasonably assured through the careful construction of a survey manual (rf. Appendix B and D). But a survey manual also serves another important function: it standardizes the conditions of administration, which is essential to reduce random error (Borg & Gall, 1983, p. 302). Although the proposed scale is generally designed for group use, where standardized conditions are not as critical as in individually administered scales (Borg & Gall, p. 302), standardized conditions are still important, the survey manual then a critical component of the prototype and the finished scale.

For the purposes of this study, the manual was addressed to the classroom teacher (although the format is appropriate for anyone administering the scale). The purpose of this was twofold: first, the classroom teacher has the ability to gain the cooperation of his/her students, and students were the target population for this study; second, the regular classroom was "probably the most desirable group to test because the students ... [were] in a familiar environment and the group ... [was] small enough so that the ... [teacher was able to] maintain good control over the situation" (Borg & Gall, 1983, p.

305; rf. also Portune, 1965, p. 42).

Communication value is another important issue, especially since the proposed scale is self-administered. The questions here are whether or not the subjects understand what they are to do and how they are to do it. To ensure such understanding, great care must be given to administration (again, the importance of the manual), directions included on the scale itself, and readability of the directions. The issue of administration with respect to the survey manual has already been addressed; however, the directions included on the scale are also a part of the administration process, and are critical with respect to communication value. Page one of the prototype (rf. Appendix C), and the finished scale (rf. Appendix E), includes the directions, which were developed in accordance with the recommendations of Anderson (1981, p. 95), and critically compared with other similar scales (rf. Chapman, 1960; Ellis, 1973; and Portune, 1965).

The administrator (i.e., the teacher) was instructed by the survey manual and the directions on the scale to read the directions aloud to the students (rf. Chapman, 1960, p. 24; and Portune, 1965, p. 43); this was especially important for the elementary students. Although no proof exists that in so doing communication value was demonstrably improved (rf. Anderson, 1981, p. 103), it did provide an opportunity for clarification. Clarification not only provides an opportunity for ensuring that the subjects understand what they are to do and how they are to do it, but also provides for an opportunity to explain the

purpose of the scale and to reinforce the psychological object of interest.

Readability of the directions is crucial for understanding (rf. Chapter V); therefore, the directions were written with the grade four student in mind, readability level theoretically established at grade four by using the formulae of Dale and Chall (1948; Instructions, 1948), Fry (1968; 1977), Rowls (rf. Anderson, 1981, p. 99), and the word lists of Dale (Dale & Chall, Instructions, pp. 45-54), and Allington et al (1985). A number of methods were used so that a "comparison of estimates" could be made, providing for increased confidence in the estimated readability levels (Anderson, p. 101).

In general terms, then, the purpose of test running the prototype (which, by the way, might be conducted a number of times before the objectives are met) was to complete the refining process with respect to the opinion statements, final decisions also made regarding administration, communication value, and scale format. The result was a finished scale, but for which no formal evaluation existed. Therefore, the next step, and the last for the purposes of this study, was to gain access to a new sample of the target group in order to estimate the reliability of the proposed scale (rf. Edwards, 1957, pp. 155-156), a step that also supplied data that were used to make a preliminary, empirical assessment of validity

Reliability and Validity. In order to obtain the data necessary to estimate reliability (and, indirectly, to assess

validity), arrangements were again made to conduct the necessary research in the school district of Abbotsford, British Columbia. As in the test run of the prototype, administration of the finished scale (rf. Appendix E), standardized by means of the survey manual developed specifically for this study (rf. Appendix D), was conducted in a regular classroom setting by volunteer teachers, who, in consultation with their school principals, were responsible for selecting the appropriate classes. An important point is that the purpose of this stage of the study was not to assess the attitudes of individual student respondents, or even of the classes, but to estimate the reliability of the proposed attitude scale. However, in order to facilitate a preliminary assessment of validity, the sample was divided into elementary and secondary groups in order to compare attitudes between the two groups. Notwithstanding, the students' right to confidentiality was respected, anonymity being guaranteed, of which they were apprised. In addition, the students were fully informed of the nature and purpose of their participation, as were their parents/guardians.

Because of practical limitations, this study depended on measures of internal consistency (split-half technique, and Cronbach's coefficient alpha) to estimate the reliability of the proposed scale. Validity was assessed by means of a logical analysis, addressing face validity and content validity, and an empirical analysis, addressing predictive validity (where the attitudes of the elementary group were expected to be more positive than the attitudes of the secondary group).

Limitations

Limitations of the study are inherent, the following points being important qualifying factors within which the study must be interpreted. First, with respect to the interviews, an essential component in the research on readability, the sample of students was not random. The interviews were confined to one elementary school and one junior high school within the same school district, and only four students each from grades four, six, and eight were interviewed. Although it is argued that the students were "average," implying that they were representative of students in the lower mainland of British Columbia, this argument is limited, restricted by the briefness of the student profiles and the size of the sample. Second, no evidence has been adduced regarding the competency and ability of the writer to conduct such interviews (a brief resume would include two years teaching experience at both elementary and secondary levels, nine years policing experience, which includes extensive work in the development of police-school liaison programs, and graduate work in education). Third, the interviews were limited by the particular interview guide used, for which no pilot study, or assessment of reliability or validity (beyond face validity), was done. Generalizations, then, beyond the students interviewed are tenuous, and all conclusions with respect to the interviews should be viewed within such a perspective.

With respect to item analysis, reliability, and validity research, similar limitations apply. For example, again, the samples were not random, the research confined to two

elementary, two junior high, and two senior high schools within the same school district. Although it is argued, as in the readability research, that these students were "average," this argument is limited for reasons similar to that in the readability research, although the samples were sufficiently large if indeed they were representative of the target population. In addition, each grade was represented by only one class of students (approximately 25 to 30), for each study, necessitating the combining of all classes into a single group for the purpose of gathering data, although sample size did permit dividing this principal group into elementary and secondary groups. Finally, no research control, apart from initial instruction to the principals involved and that provided by the survey manuals, existed over the volunteer teachers; and no evidence is adduced that the competency and ability of the teachers were equal.

Additional limitations apply equally to all the research conducted in this study. For example, all research is limited temporally: readability limited to early 1987, item analysis to late 1987, and reliability and validity to early 1988. In addition, no explanation into the prevailing conditions of the time, which may have influenced the way in which students responded to the proposed scale, is offered.

This study is also limited, of course, by its definition of attitude, and by assumptions inherent within the quantitative approach to attitude assessment, such as the assumption that a person's attitude can be transformed into a number and fixed

along a quantitative continuum, and that this will provide a reasonably accurate assessment of his/her attitude regarding some particular object of interest (rf. Eisner, 1983, & 1985; and Smith, 1983). Finally, this study is limited by the research capabilities and descriptive interpretations of the writer. In conclusion, then, generalizations beyond the classes and schools involved and the times described are limited by this study's overall methodology; but in addition, generalizations within the classes and schools involved and the times described are limited by assumptions inherent in quantitative methodology, which some argue is unique to the epistemology of the "realist" (rf. Smith).

CHAPTER IV

Preliminary Instrument Construction

This chapter details the preliminary steps that were followed in the construction of the proposed scale: (1) defining the theoretical foundation; (2) assembling an item pool; and (3) making a preliminary statement selection from the item pool, based on the defined theoretical foundation and accepted criteria respecting summated ratings scales as defined in the literature.

Theoretical Foundation

This study was concerned only with student attitudes toward the police, impressions of which described along a unidimensional continuum. It is true that the police may be symbolic of the law, that external differences between the police exist (e.g., federal, provincial, city, etc.), and that differences within the police exist (e.g., uniformed patrol, plain clothed detective, traffic, etc.), but research indicates that these differences are not significant with respect to attitudes toward the police in general (rf. Koenig, 1974, & 1975; and Phillips & Coates, 1971).

Here, then, it is important to understand how the police are perceived by citizens of a democratic society. In this study's introduction, the point was briefly made that notwithstanding legislative authority, in a democracy the ability of the police to perform their duties is dependent upon

public approval, support, and willing cooperation. In a police state, coercion is the standard, but in a democracy, cooperation is the standard. In other words, the police in a democratic society function under a unique and distinct mandate, which is ultimately authorized by the public. In a discussion of this unique and distinct mandate, this chapter will briefly review the historical roots of policing in England, Canada, and the United States, examine the role of policing in a democratic society, and define the theory on which policing in a modern democratic society, such as Canada, is based.

People have always had some form of law and system of law enforcement, regardless of the society. People need rules to survive as a society; subsequently, total freedom is an abstraction that has no practical meaning. Modern law, as it is found in the United States and Canada, developed "directly from the system of law in England" (Beckman, 1980, p. 1; rf. also Loree, 1985; Salhaney, 1984; Shane, 1980; and St. Pierre, 1981). In 1829 Sir Robert Peel, a member of English Parliament, proposed a bill that would form the foundation for policing in England and many other countries of the world. His suggestions were accepted by Parliament and became operationally effective on September 29, 1829 (Beckman, p. 30). In his proposal, Peel outlined 12 principles for effective policing consistent with a democratic society. These 12 principles have been condensed, and today are traditionally known as Peel's nine principles (rf. Beckman, p. 30; and Radelet, 1973, pp. 4-5).

Peel based his principles on the belief that it is the

responsibility of every citizen of a democracy to actively participate in the preservation of law and order, the police simply an arm of the citizen; i.e., "the police are the public and ... the public are the police" (rf. Radelet, 1973, pp. 4-5). Peel's principles were revolutionary, and became an inherent part of the evolution of democracy, significantly influencing policing in the United States, India, Israel, Netherlands, and Canada (Shane, 1980).

However, the role of the police, notwithstanding its roots, is "one of the central, most perplexing, most fundamental questions..." of a democratic society, based on "the ancient riddle of freedom and order" (Radelet, 1973, pp. 32, 41). Because of the critical role that the police play in a democratic society, this question has attracted a great deal of attention by scholars, politicians, police, and other interested people. For example, Beckman (1980) states, "The most significant characteristic of the unique role the policeman plays in the community is that very few people actually know what that role is" (p. 88); Jaywardene (1973) states, "The police role has become an almost undefinable concept" (p. 14); Shane (1980) describes the police role as "ambiguous" (p. 4); and Loree (1985) states, "There is no universal consensus..." of the police role (p. 410).

An illustration of "the ancient riddle of freedom and order" can be seen in the following example of extreme views. Pfiffner (1969) believes "the point to be made ... is that the police subculture still possess a set of values, standards and

job goals more appropriate to the days of hangings than to a society which is making some progress toward ameliorating the lot of those who ... have not adjusted to the demands of society" (p. 6). On the other hand, Bewley (1986) believes that academics and legislators "took a country in which the citizens never before needed to lock up their homes and possessions and could safely walk anywhere, anytime, unafraid; they turned it into a country where the thugs increasingly freely roam the streets and the citizens instead are increasingly prisoners in their own homes" (p. 5).

As a result of this confusion, conflicting demands are placed on the police. Those with a legal emphasis see the police role primarily as law enforcer, crime fighter, controller, and state official; however, those with a social emphasis see the police role primarily as peace officer, crime prevention officer, supporter, and community citizen. In addition, many special interest groups, each with their own agenda, attempt to influence the police to suit their own needs. In extreme cases, police action may be deemed illegitimate by significant groups of society because of their perceptions of the police role. For example, those striving for radical social change often perceive the police to be "representatives of the 'ruling' class and oppressors, even often as the enemy itself" (Shane, 1980, p. 2).

Because of the complex, sophisticated society in which we live, it is inevitable that the role of the police will be questioned and debated. This, however, is not to be

discouraged, but is part of the democratic process in search of the ideal. There is no absolute answer to the "ancient riddle"; notwithstanding, the issue is resolved in the will of the people, which is fundamental to a democratic society.

Despite the controversy respecting the role of the police in a democracy, general consensus is found in the roots of policing, where one finds concepts that remain accepted to this day. Lawlessness cannot be tolerated in a democratic society which upholds the "rule of law"; freedom without form is anarchy. If the need for law is a given, law enforcement must follow, for laws without the means to enforce them are meaningless. The questions, then, are what laws and how are they to be enforced? These questions, in a democracy, must be answered by the will of the people. According to Gall (1983), our justice system depends "upon the acceptance of the system by those it governs in order to achieve its desired objectives. A legal system must command the support of the members of society, for without general social acceptance it simply cannot function, at least not in the context of a liberal, democratic society [underlining added]" (p. 2).

Sir Robert Peel articulated this precept back in 1829, whose message was one of cooperation rather than coercion, a message which was to become a major philosophic influence in the development of policing in a democratic England (Shane, 1980, p. 13), and which subsequently permeated the whole concept of justice in modern, democratic societies. What Peel recognized was that the strength of the police was to come from a moral

position, not physical power (Shane, p. 14); a recognition still echoed today. Loree (1985), for example, states that in Canada, everyday policing occurs by consent "and is thus dependent on the moral rather than the physical authority of the police" (p. 407).

Peel, then, defined policing in a democratic society. A knowledge of his principles is a prerequisite for any understanding of policing in England, the United States, and Canada, and are summarized below.

The responsibility of the police is:

1. To prevent crime and disorder.
2. To recognize that policing authority is dependent upon public approval and cooperation.
3. To secure and maintain the respect and approval of the public.
4. To recognize that public cooperation is directly related to the amount of physical force needed to achieve social order.
5. To seek and preserve public favor by being sensitive to public opinion, by being courteous and friendly, and by being impartial.
6. To use a minimum of force, and only when necessary.
7. "To maintain at all times a relationship with the public that gives reality to the historic tradition that the police are the public and that the public are the police; the police being only members of the public who are paid to give full-time attention to

duties which are incumbent, on every citizen, in the interest of community welfare and existence" (rf. Radelet, 1973, pp. 4-5).

8. To refrain from usurping the power of the judiciary.
9. To recognize that the test of police efficiency is the absence of crime and disorder.

In conclusion, then, I would quote Justice Patrick Hartt, former chair of the Law Reform Commission of Canada, as found in Gall (1983):

> What we are really talking about is a society in which citizens are not simply joined collectively together by the threat of a coercive state, but rather one in which they are united from within themselves, by agreement on core values and a shared desire to secure their common goals. (p. 324).

The Item Pool

That the principles of Sir Robert Peel are a part of our heritage is an historic fact, and that they are at least a logical starting point for the selection and judgment of opinion statements for the proposed attitude scale is demonstrated by concerns of the public today, evidenced daily in newspapers, television, radio, civil rights publications, etc. For example, current topics include concern with police effectiveness in controlling crime, corruption, impartiality, abuse of authority, and use of force (compare the following headlines which all appeared in the Vancouver Sun newspaper: Emphasis on

high-technology police enforcement deplored (1987, August 8); Police facing brutality probe (1987, May 1); System keeps B.C. police neutral in labor disputes (1987, September 9); Victim of thefts unhappy with police (1987, March 9); Watching the watchdogs (1986, December 8)). These concerns illustrate that Peel's principles are still applicable today; and, therefore, formed the basis for judging whether or not opinion statements adequately reflected the domain of interest.

A pool of opinion statements, 79 in total, was obtained from a number of studies (Chapman, 1960; Ellis, 1973; Pegler & Wright, n.d.; Phillips & Coates, 1971; Portune, 1965) that included sentence based scales constructed to assess a common psychological object of interest, attitudes toward the police (rf. Hess & Torney, 1967, p. 233). Of these statements, 40 were eliminated, identified as statements that did not meet the theoretical foundation as outlined in this chapter, or did not meet the accepted criteria for inclusion in a summated ratings scale as defined in the literature (rf. Anderson, 1981, pp. 248-250; Borg & Gall, 1983, p. 421; Edwards, 1957, pp. 13-14; Girod 1973, Ch. 4; Gronlund, 1985, p. 419; and Likert, 1932, pp. 44-53). These criteria are summarized in the following points.

1. Eliminate redundant or repetitive statements.
2. Imperative to measure a present attitude.
3. Avoid statements of fact.
4. Avoid irrelevant statements.
5. Must be clearly negative or positive with respect to the psychological object.

6. Avoid statements that all or none will endorse.
7. Avoid ambiguous statements.
8. Avoid using words or jargon that may not be uniformly understood.
9. Statements should be short, rarely exceeding twenty words.
10. Each statement should have one, clear thought.
11. Statements should be in the simplest possible vocabulary.
12. Avoid compound sentences.
13. Absolutely no double negatives.
14. Use negatives sparingly, and when used, underline them.
15. Eliminate statements that are temporally or culturally restrictive.
16. Eliminate or rewrite sexist statements.
17. Both extremes of the attitude continuum should be equally utilized.
18. Statements should be randomly distributed.

Note that some authors, such as Edwards (1957), recommend against using a complex sentence, believing it unnecessarily complicates the statement. However. research has indicated that in some instances the complex sentence enhances comprehension, and that there is no evidence to suggest that periodic sentences are any more difficult than loose sentences (Fry, 1977, p. 250).

Preliminary Statement Selection

The 39 remaining statements were again scrutinized, and 19 statements that appeared to fulfill best the described criteria and to reflect best the domain of interest were selected. These statements were as follows.

1. Police represent trouble instead of help.
2. The police stop and question people because it makes them feel powerful.
3. The police are not needed in our community.
4. Police enjoy kicking people around.
5. On the whole, policemen are honest.
6. The police are stupid.
7. Police often overstep their legal authority.
8. Policemen are just as crooked as the people they arrest.
9. Police are brave men.
10. The police are always mad at kids.
11. Police officers care for the younger generation.
12. The police are doing a good job.
13. Policemen don't care what happens to you after they pick you up.
14. If I ever get into trouble with the police, I could buy my way out.
15. Policemen are a great help to folks.
16. Young people are treated fairly by police officers.
17. Police officers are friendly and approachable.
18. Policemen are pretty nice guys.

19. Without policemen there would be crime everywhere.

At this point, it is important to note that these statements were in an unfinished form. For example, some were sexist, inconsistencies existed in how the police were designated (e.g., the police, police, policemen, police officers), and readability had not yet been addressed.

CHAPTER V

Readability

Following the selection of 19 opinion statements, it was necessary to examine the issues of readability and ability of elementary students to differentiate between the responses of the standard five-point summated ratings scale. Once these issues had been satisfactorily addressed, a selection of statements appropriate for item analysis was made. The design of this research included two major stages: first, the issues were resolved on a theoretical level; second, in order to operationally assess the theoretical conclusions, personal interviews were held with students in grades four, six, and eight.

The Theory

Readability is usually based on the difficulty level of the vocabulary and the complexity of the sentence structure (Anderson, 1981, p. 97; Fry, 1977, p. 244), and may be determined in a number of ways. For example, readability level may be established through formulae based on word lists, formulae based on complexity of sentence structure, a combination of these types of formulae, oral reading, personal interviews, expert judgment, comparison of written passages with material of known traditional difficulty, and other more obscure techniques (Fry, p. 248). However, despite subjective opinions on which method is best, objective evidence indicates that most

formulae and methods correlate "quite highly" with each other (Fry, 1968, p. 516); and Fry, quoting G. R. Klare, "a widely recognized bibliographer of readability studies," points out that "there is little to be gained from choosing a highly complex formula" (1977, p. 244).

Given the pencil and paper nature of self report instruments, and the fact that the Likert technique is based on a series of sentences, the factor of readability was critical in the construction of the proposed attitude scale because the target population consisted of students in grades as low as four. As the pool of statements collected for this study was obtained from scales intended for junior and senior high students and adults, it was necessary to reduce the readability level to that of grade four. Of necessity, most statements required rewriting, although often of a minor nature; and readability level was established (rf. Table 5-1) using the formulae of Dale and Chall (1948; Instructions, 1948), Fry (1968; 1977), Rowls (rf. Anderson, 1981, p. 99), and the word lists of Dale (Dale & Chall, Instructions, pp. 45-54), and Allington et al (1985).

The formula of Dale and Chall (1948; Instructions, 1948) was used by (1) determining the average sentence length of the entire sample of words by dividing the number of words by the number of sentences, (2) determining the percentage of unfamiliar words in the sample of words by using the Dale List of 3000 Familiar Words (Instructions, 1948), and (3) inserting this information into their formula and subsequently estimating

readability level.

Table 5-1
Readability Level

Source	Level
Dale & Chall	Gr. 4.2
Fry	Gr. 4.0
Rowls	Gr. 4.0

The formula of Fry (1968; 1977) involved (1) determining the mean number of sentences from three 100 word samples, (2) determining the mean number of syllables from the three 100 word samples, and (3) plotting these mean scores on Fry's graph, which estimates readability level. Rowls's (Anderson, 1981, p. 99) formula involved using the first two steps described for Dale and Chall, and subsequently using this information to estimate readability from a chart developed by Rowls. The word list of Allington et al (1985) provided a check for the word list of Dale, which is approximately 40 years old. The purpose of using a number of methods to estimate readability was to provide a "comparison of estimates," which allowed for increased confidence in the estimated readability level (Anderson, p. 101).

Because of the nature of the standard summated ratings scale with its five response categories, assessment of readability also included the issue of the ability of elementary students to differentiate between the responses. First, the responses selected were "strongly agree, agree, unsure,

disagree, strongly disagree" because they were appropriate and because they appeared in the Dale List of 3000 Familiar Words (Dale & Chall, Instructions, pp. 45-54) and Allington's et al list of frequently appearing words (pp. T42-T49), both of which include all familiar words at the grade four level.

Second, the format of the responses had to be resolved; for example, using the symbols "SA A U D SD" vis-a-vis the words "Strongly Agree, Agree, Unsure, Disagree, Strongly Disagree." Here, a decision was made in favor of writing out the responses in words for the benefit of the younger students; as Likert (1932) stated, "It is a desirable precaution to state each proposition in such a way that persons of less understanding than any member of the group for which the test is being constructed will understand and be able to respond to the statements" (p. 45). Likert, in his study, "A technique for the measurement of attitudes," wrote out the responses in words, although he used the root words "approve" and "decide" ("approve" was not an appropriate response, and "decide," which already had two syllables, required both a prefix and a suffix for the response "undecided," whereas the response "unsure" required only a prefix, making it an easier word to read than "undecided").

Finally, taking another "precaution" in favor of ease of readability, the response format selected was to place the responses in multiple-choice fashion (Gronlund, 1985, p. 419), the following an example:

The police enjoy kicking people around.

( ) Strongly Agree

( ) Agree

( ) Unsure

( ) Disagree

( ) Strongly Disagree

In conclusion, as a result of rewriting the statements, and as a result of writing out the responses in a multiple-choice format, the original 19 opinion statements were revised as follows, order determined by random assignment (the responses included for the first statement only, for illustrative purposes).

1. The police are trouble instead of help.

   ( ) Strongly Agree

   ( ) Agree

   ( ) Unsure

   ( ) Disagree

   ( ) Strongly Disagree

2. The police stop and question people because it makes them feel powerful.
3. The police are not needed in our country.
4. The police enjoy kicking people around.
5. On the whole, the police are honest.
6. The police are dumb.
7. The police break the law.
8. The police are just as crooked as the people they put in jail.
9. The police are brave.

10. The police are always mad at kids.
11. The police care for young people.
12. The police are doing a good job.
13. The police don't care what happens to you after they pick you up.
14. If I ever get into trouble with the police, I could buy my way out.
15. The police are a great help to people.
16. Young people are treated fairly by the police.
17. The police are friendly.
18. The police are pretty nice people.
19. Without the police, there would be crime everywhere.

The Interviews

Although the revised statements were theoretically appropriate for students in grades as low as four, a critical question was whether or not elementary students could differentiate between the responses as required by the standard summated ratings scale. Numerous examples of published tests that use the five-point response for children as young as those found in grade four can be found (rf. Johnson, 1976, Category 8: Attitudes and Interests). Alternatively, though, scales such as the Estes Attitude Scales use a modified three-point response for elementary students, and the standard five-point response for secondary students (Gronlund, 1985, p. 420).

A related issue here is whether or not the proposed attitude scale is appropriate for the range of grades for which

it is intended. At this point, it is sufficient to point out that numerous examples of scales that are intended for a wide range of grades exist (rf. Finch & Rogers, 1984), and Anderson (1981) states that "if the empirical evidence remains fairly constant across the grade levels, this would suggest that concerns ... are not warranted" (p. 122). Subsequent "empirical" research addressed this issue by means of statistical analyses (e.g., as found in item analysis and reliability).

In order to operationally assess the theoretical conclusions respecting readability and differentiation by elementary students on the responses, a decision was make to conduct personal interviews with a small but representative sample of students in grades four, six, and eight, such a tactic recommended in the literature. For example, Chapman (1960) and Hess and Torney (1967), when constructing their attitude scales to assess the attitudes of children and adolescents, conducted preliminary interviews for the purpose of assessing readability. Anderson (1981), in a discussion of readability, suggests tryouts; and Fry (1977), for final confirmation, suggests that "there is no substitute for trying out the passage ... on a sample of population for whom ... [it] is intended" (p. 246).

The objectives of the interviews were (1) to assess the readability level of the revised opinion statements , the assessment based on oral reading, vocabulary, comprehension, and predictions by students, (2) to determine whether or not students at the elementary level (the lowest grade being four)

have the ability to differentiate between the responses as required on a five-point summated ratings scale, and (3) to identify statements that would be most appropriate for item analysis.

The interview is a valid method of data collection in the social sciences (Brenner, Brown, & Canter, 1985, p. 1), and it has a number of advantages over the pencil and paper questionnaire. For example, it is adaptable and flexible, which makes it "especially suitable for children" (Kerlinger, 1964, p. 467). The interview "allows both parties to explore the meaning of the questions and the answers involved," and misunderstandings can be immediately identified (Brenner et al, p. 3).

Paradoxically, though, the unique strengths of the interview, flexibility and adaptability, allow for its greatest weaknesses, subjectivity and bias (Borg & Gall, 1983, p. 437; Brenner et al, 1985, pp. 2, 4; Parker, 1984, pp. 19, 21). For example, bias may be found in the interviewer, who may be seeking out answers to fit a preconceived idea, and "response effect" may be found in the respondent, who may be eager to please. In addition, other weaknesses of the interview include the time factor involved in conducting individual interviews, and the accuracy of the respondent, who may be lying, withholding the truth, and/or unable to answer and so fakes a response (Parker, p. 22).

For these reasons, the interview must be justified over the simple pencil and paper questionnaire (Borg & Gall, 1983, p.

436; Parker, 1984, p. 27), such justification for this study found in the objectives herein presented. For example, personal interviews were necessary for the oral reading component in assessing readability, and for the role playing component in determining ability to differentiate on the five-point response.

The Student Interview Guide. Having justified the interview, and in order to deal effectively with the inherent weaknesses of the interview and to facilitate the accurate collection of data, it was imperative to design an interview guide (rf. Borg & Gall, 1983, pp. 441-442; Kerlinger, 1964, p. 467; and Parker, 1984, p. 27). Therefore, the Student Interview Guide (rf. Appendix A) was constructed according to accepted criteria as defined in the literature (rf. Baker, 1983, p. 509; Borg & Gall, p. 443; Farley, 1981, p. 185; Gross, 1984, pp. 67-68; Kerlinger, p. 474; Parker, pp. 20-21, 27; and Simmons, 1974, Ch. 2). These criteria are summarized as follows.

1. Both the interview and the environment should be standardized.
2. Define objectives.
3. Must be designed with the verbal and cognitive abilities of the respondents in mind.
4. Define format, wording, allowable probes, recording procedures, and time frame.
5. Allow ample time for interview.
6. Address the issue of respondent cooperation.
7. Assure confidentiality.

8. Respondent should be fully informed, unless it is appropriate not to do so.

9. Plan introduction, setting respondent at ease.

10. Avoid leading questions.

11. Avoid threatening questions.

12. Each question should be focused on one idea.

13. If appropriate, let respondent represent peer group, to facilitate cooperation and accuracy.

14. Plan for post-interview equilibrium.

Results and Discussion.

Table 5-2
Readability of Statements

	Type:	Oral Reading			Vocabulary			Comprehension			Prediction		
	Gr:	4	6	8	4	6	8	4	6	8	4	6	8
	1	X						X		XX			
	2							X					
	3												
	4												
S	5							XX	XX	X	X	X	
t	6												
a	7												
t	8	X	X	X	XX	XX	X				X	X	X
e	9												
m	10												
e	11												
n	12												
t	13												
s	14							XX	XXX	XX	X	XXX	
	15												
	16												
	17												
	18												
	19												

X: a student who had a problem or predicted a problem.

The first objective of the interviews was to assess the readability level of the statements, which had been theoretically established at grade four. Table 5-2 shows the results of this part of the interviews, indicating readability problems with statements 1, 5, 8, and 14. Note that problems with these statements were not confined to any particular grade, occurring fairly consistently across grades. Note also that where a problem tended to occur, students would fairly consistently predict that their peers would have difficulty with the statement in question, verifying the assessment.

Statement one read, "The police are trouble instead of help." One student had a problem here with oral reading, involving the word "instead," and three students had problems with comprehension, involving the statement's position. For example, nine students interpreted the statement as one being negative about the police (the original intent of the statement), but three students interpreted the statement as one being positive about the police, believing it to mean that the police represent trouble to criminals. Because of these problems, this statement was eliminated from the item pool.

Statement five read, "On the whole, the police are honest." Five students had problems here with comprehension, including three predictions of problems. The source of difficulty was simply with the prepositional phrase, "on the whole," which these students just did not understand, although they were able to correctly guess the meaning of the statement when prompted to do so. Rather than eliminating this statement,

then, the prepositional phrase in question was dropped, retaining the simple, independent clause, "the police are honest," which was clearly understood by all students.

Statement eight read, "The police are just as crooked as the people they put in jail." Three students had problems here with oral reading, five with vocabulary, including three predictions of problems. The single problem here was with the use of jargon, that being the term "crooked," which reaffirmed the rule against the use of jargon. However, notwithstanding this problem, all students were able to correctly guess the meaning of the statement when prompted to do so, often substituting the word "bad" for "crooked." For this reason, the word "crooked" was replaced by the word "bad," the statement then retained.

The last statement where problems frequently occurred was number 14, which read, "If I ever get into trouble with the police, I could buy my way out." This statement proved to be most troublesome, with seven students having problems with comprehension, including four predictions of problems. The source of the problem was with the clause, "I could buy my way out." This again reaffirmed the rule against using jargon--"buy my way out" was intended to refer to a bribe, a negative statement about the police, but students often interpreted it as referring to "bail," becoming more a statement of fact (an interesting interpretation). Because this problem could not easily be corrected, this statement was eliminated from the item pool.

Continuing with the first objective, the readability of the responses (strongly agree, agree, unsure, disagree, strongly disagree) was assessed. No problems were identified in any of these areas, including the response "unsure," making it unnecessary to deal with alternatives for this response (rf. Student Interview Guide, p. 12).

The second objective of the interviews was to determine whether or not students at the elementary level (the lowest being grade four) have the ability to differentiate between the responses as required on a five-point summated ratings scale. The theory here was that if the students could select a statement, assume a role randomly assigned, read the statement silently, and subsequently respond to the statement in a manner consistent with the role, then one could reasonably assume that elementary students could differentiate between the five responses. The test here was rigorous, and did not cater to the younger student--consider that the student was alone in a room with a stranger, probably somewhat apprehensive, and was asked to set aside his/her attitudes, assume a role, correctly interpret a statement within the context of the role, and then respond in a manner consistent with the role. Not only did such a test assess ability to differentiate on the five-point response, but it also reassessed comprehension, as it required the student to understand what was read and to make a judgment about what was read (Allington et al, 1985, p. T32). In addition, then, this exercise indirectly assessed the content of the statements; i.e., whether or not the content was appropriate

for the target population.

Table 5-3
Ability to Differentiate Between Responses

	Grades		
Error type	4	6	8
Direction	XXX	X	XX
Degree			

X represents a mistake

Table 5-3 shows the results of this part of the interviews, indicating that students in grades four and six did not make more mistakes than those students in grade eight, this evidence supporting the conclusion that elementary students in grades as low as four have the ability to differentiate between the responses as required on a five-point summated ratings scale (cp. Hess & Torney, 1967, p. 226). As the results show, out of sixty roles played (each student in each grade played five different roles), three mistakes were made in grade four, (two mistakes were made by the same person), one mistake was made in grade six, and two mistakes were made in grade eight; or in terms of a percentage score for each grade as a whole, grades four, six, and eight scored 85%, 95%, and 90% respectively.

A final point for consideration, which corroborates the conclusion that elementary students have the ability to differentiate between the responses, is the fact that all the mistakes were made in direction of response, not in degree of

response. The suggestion here is that the students who made mistakes may have been confused in their role or misunderstood the statement with respect to whether they were to respond positively or negatively; this is an understandable mistake, considering the demands made on the students.

A point of caution regarding this conclusion, however, is the qualifying factor that in addition to the stated limitations of the study, the role playing game was limited by its subject matter; i.e, student attitudes toward the police. Although the evidence suggests that students in grades as low as four have attitudes about the police that are sufficiently formed to enable them to differentiate between the five-point response of the standard summated ratings scale, the evidence does not suggest that these students have the universal ability to differentiate between the five-point response of the standard summated ratings scale, regardless of the subject matter.

Conclusions

As a result of revising the 19 opinion statements selected from the item pool, designation of the police was made uniform (i.e, "the police," basically selected because it is not sexist and it is a familiar word for students in grade four), and readability was theoretically established at grade four. As a result of the interviews, statements 1 and 14 were eliminated, statement 5 was altered by dropping the prepositional phrase "on the whole," statement 8 was reworded by replacing the word "crooked" with the word "bad," and the response "unsure" was

found to be acceptable. In addition, it was determined that students in grades as low as four were able to respond to the standard five-point summated ratings scale. Therefore, based on theoretical and practical considerations, 17 statements, along with the standard five-point response written out in multiple-choice format, were found to be appropriate for item analysis (rf. Appendix C).

CHAPTER VI

Item Analysis

The extent to which the revised 17 statements were related and assumed to be a scale was determined through item analysis, a process which took two forms, logical analysis and empirical analysis, both complementary to each other (rf. Anderson, 1981; and Gardner, 1975). Logical analysis refers to qualitative evidence, focusing on the construction, rationale, and foundation of the statements, whereas empirical analysis refers to quantitative evidence, focusing on the responses made to the statements by a sample group. This chapter, then, reports the results, first, of a logical analysis, and second, of an empirical analysis.

Logical Item Analysis

Without getting into a debate on the superiority of any particular method of analysis, which actually becomes quite philosophical (rf. Smith, 1983), statements must first stand up to logical analysis; otherwise, statistical analysis may be quite absurd. According to Gardner (1975), there should always be a "clear attempt to relate ... [item] content to an explicit rationale" (p. 106). This "clear attempt," in effect, forms a defendable philosophy, and it is to this that this study now turns.

For this study, the philosophical principles of Sir Robert Peel formed a logical basis, preceding statistical analysis, to

select statements which appeared to be related to each other and which appeared to reflect adequately the domain of interest. Each of the 17 statements identified for item analysis were scrutinized for consistency with the described theoretical foundation, and appeared to be related to each other, justifying their treatment as a unit with respect to the psychological object of interest (rf. Likert, 1932, p. 23). Although this logical analysis did not suggest that any particular statement was inappropriate insofar as the theoretical foundation was concerned, two statements did appear to be suspect because they did not appear to meet the accepted criteria for inclusion in a Likert-type scale as defined in the literature.

The first suspect statement was number three, "The police break the law." I originally had concerns about this statement because everyone breaks the law to some degree, and so the statement seemed to be one of fact rather than attitude; in addition, it was rather ambiguous: did "break the law" refer to serious offenses, corruption, or the minor breaches that all commit from time to time? For these reasons, it did not appear to be a good statement; i.e., one that would consistently discriminate between persons of differing attitudes.

The second suspect statement was number six, "Without the police, there would be crime everywhere." Again, as with statement number three, this seemed to be more a statement of fact rather than attitude. It seemed reasonable to assume that only the most idealistically naive would believe that crime would not increase if the policing component of the criminal

justice system were eliminated. In other words, most people, whether their attitudes are positive or negative towards the police, would agree with this statement, and the statement therefore would fail to discriminate between persons of differing attitudes.

Logical item analysis, therefore, supported the inclusion in a finished scale of all 17 items with respect to the theoretical foundation; however, the analysis did bring into question two statements that did not appear to meet the accepted criteria for inclusion in a summated ratings scale.

Empirical Item Analysis

A summated ratings scale demands that each statement be related to each other, and that each statement be discriminating; i.e., that differing attitudes are consistently reflected in the responses (Adams, 1982, p. 181; Edwards, 1957, pp. 12, 154; Edwards & Kenney, 1967, p. 250; Lemon, 1973, p. 180; Likert, 1932, pp. 23-48; Portune, 1965, p. 25; Rundquist & Sletto, 1975, pp. 11, 306). A number of different methods have been developed for these purposes, some of the most common being Likert's "criterion of internal consistency" (actually, Likert was not the first to use this method, which was used as early as 1916 in the Stanford-Binet Scale (Rundquist & Sletto, p. 4)), correlational procedures, and t-tests. The results of these methods are often quite similar, and so it is sometimes recommended to use the simpler method, the criterion of internal consistency (Edwards, pp. 152-155; Likert, p. 50). More

recently, however, correlational procedures are ordinarily used (Scott, 1968, p. 219), although some researchers (rf. Gardner, 1975; and Triandis, 1971), criticizing the almost exclusive use of sophisticated statistical techniques as a basis for decision making, continue to develop scales by relying on well developed theory and logical analysis (Adams, p. 181).

Empirical analysis, in addition to identifying statements that appear related and discriminating, also determines whether or not the numerical values are properly assigned to the responses in the case of Likert-type scales (Likert, 1932, p. 48; Rundquist & Sletto, 1975, p. 318). Finally, empirical analysis can serve still another purpose, one especially important to this study, which is to determine if the scale is appropriate for the intended grades (Anderson, 1981, p. 122).

With respect to the empirical analysis in this study, a correlational method was used as the primary source of evidence upon which to make decisions concerning the statements. In addition, though, the criterion of internal consistency was used as a secondary check, providing for increased confidence in determining the best statements. The particular correlational procedure used was relatively simple: a correlation coefficient (Pearson's product moment correlation) was calculated for each statement, where the response score was correlated with the battery score for each subject (rf. Edwards, 1957, p. 155; Lemon, 1973, p. 180; Likert, 1932, p. 48; Phillips & Coates, 1971, p. 6; and Scott, 1968, p. 219). The problem here, though, was that because the battery score included the response score

the results were artificially inflated (rf. Lemon, p. 180; and Scott, p. 219). There were two possible solutions to this problem: one was to use the Peters and Van Voorhis (1940) correction formula; the other was to simply correlate the response score with the battery score minus the response score (rf. Phillips & Coates, p. 6), and it was this method which was used in this study.

The size of the correlation coefficient, or the critical ratio as it applies to the criterion of internal consistency, between any particular statement and the battery justifying inclusion of the statement in the finished scale is entirely arbitrary, based on the researcher's judgment (Edwards & Kenney, 1967, p. 254; Likert, 1932, p. 49; Rundquist & Sletto, 1975, pp. 11, 16), which takes us back to logical analysis and the importance of well developed theory. For the purposes of this study, statistical data were used as a guide, as corroborating evidence, and it is to these data that this study now turns.

The data in Tables 6-1 and 6-2 show that the sample was representative of the group of interest; however, sample size restricted an empirical item analysis to the group as a whole and to the elementary and secondary groups, as previously discussed in Chapter III. Because the size of the total sample exceeded the minimum requirements as defined in the literature, it was used as the primary source of data upon which to make decisions concerning statement selection. In addition, because the sample size of the elementary and secondary groups met most minimum requirements, they also were analyzed to determine if

their results were generally consistent with the total sample, and so supplied additional data upon which informed decisions were made.

Table 6-1
Statistical Description of the Sample

Grades	Mn Age	N (Male)	N (Fem)	N (Total)
4	9.1	13	10	23
5	10.0	14	6	20
6	11.0	13	17	30
7	12.0	14	24	38
8	13.3	19	7	26
9	14.5	8	18	26
10	15.0	19	11	30
11	16.1	13	9	22
12	17.3	13	17	30

Table 6-2
Statistical Summaries of the Sample

Grades	Mn Age	N (Male)	N (Fem)	N (Total)
El:4-7	10.7	54	57	111
Sec:8-12	15.3	72	62	134
Tot:4-12	13.2	126	119	245

Correlation Coefficients. Table 6-3 shows the correlation coefficients for each statement, the coefficients for statements 6 and 12 appearing relatively low. Statement six was identified in logical analysis as being suspect, a conclusion corroborated here. As stated previously, the size of the coefficient as a basis for statement selection or rejection is arbitrary; consequently, coefficients of approximately .5 or higher were considered acceptable for the purpose of this study. Similarly,

Phillips and Coates (1971) in their construction of a summated ratings scale (designed to assess student attitudes toward the police), used .6 as the cutoff criterion, their final scale showing coefficients of .5 to .8 (pp. 5-6).

Table 6-3
Correlation Coefficients

	Group		
Statements	El	Sec	Tot
1. The police are pretty nice people.	.58	.74	.74
2. The police are brave.	.53	.53	.58
3. The police break the law.	.53	.37	.55
4. The police are a great help to people.	.47	.61	.64
5. The police enjoy kicking people around.	.54	.63	.64
6. Without the police, there would be crime everywhere.	.20	.30	.25
7. The police are doing a good job.	.57	.62	.66
8. The police are dumb.	.61	.67	.66
9. The police are just as bad as the people they put in jail.	.61	.58	.64
10. The police are honest.	.53	.53	.60
11. The police are friendly.	.65	.70	.72
12. The police are not needed in our country.	.27	.48	.42
13. The police don't care what happens to you after they pick you up.	.54	.55	.62
14. The police are always mad at kids.	.52	.60	.63
15. The police care for young people.	.64	.65	.70
16. Young people are treated fairly by the police.	.54	.55	.57
17. The police stop and question people because it makes them feel powerful.	.44	.53	.54

Note: Statement position was determined by random assignment.

Statement 12 was not previously identified, but on close inspection, it appeared to be stating the same concept as statement 6. Therefore, it was only logical that both would have relatively low and similar coefficients. The evidence, then, appeared strong enough to support a rejection of both

these statements.

Upon calculating the coefficients of the elementary and secondary groups, the coefficients for statements 6 and 12 were predictably low, but the coefficients for statements 3, 4, and 17 were also relatively low. In statement three, the coefficient was only .37 for the secondary group. In the logical analysis, statement three was identified as suspect because it appeared to be ambiguous and one of fact. Although the elementary group's coefficient was .53, the item to total correlation for the secondary group was unacceptably low. In statement four, the coefficient was .47, relatively low, for the elementary group, but was over .6 for the secondary group. Logically, it was difficult to understand why this statement failed to discriminate well amongst the elementary group.

Statement 17 also had a low item to total correlation for the elementary group. This, however, was attributed to a statement where the meaning was too subtle for elementary students. In addition, in a strict sense, the statement was grammatically incorrect because the pronoun "them" technically referred to "people," which was the antecedent rather than "police," which was the intended antecedent (rf. Frank, 1972, pp. 44-46). Statements 3, 4, and 17, therefore, appeared not to be satisfactory statements.

Criterion of Internal Consistency. In addition to computing correlation coefficients for each statement, critical ratios (rf. Table 6-4) were computed using the criterion of

internal consistency. This test of internal consistency was computed by simply recording the responses of the subjects who fell into the two extremes of the attitude continuum. The scores for each statement were summed for both groups; following this, the totals for the low group were subtracted from the totals for the high group, this difference subsequently divided by the number of subjects constituting one group. The result was a critical ratio, which was used to select the most discriminating statements.

The size of the groups researchers use to represent the two extremes of the attitude continuum usually varies between 10% (that used for this study) and 25% (rf. Edwards & Kenney, 1967, p. 25; Lemon, 1973, p. 180; Likert, 1932, p. 51; and Rundquist & Sletto, 1975, p. 11), the decision being entirely arbitrary. It is important to note, however, that the larger the extreme group used, the lower will be the critical ratios (Rundquist & Sletto, p. 111). These critical ratios, then, are relative, but the ratio of 1.8 is often used as the cutoff criterion for a group of 10% (rf. Edwards & Kenney, p. 254). Again, however, the size of the critical ratios is a function of the size of the difference between the means of the extreme groups; therefore, if the sample is not diverse or varied, the critical ratio will be smaller than if a sample was purposefully selected because of great diversity.

Table 6-4 shows the critical ratios for the total sample and the elementary and secondary groups.

Table 6-4
Critical Ratios

Statements	Group El	Sec	Tot	Statements	Group El	Sec	Tot
1	1.3	2.6	2.8	10	1.6	1.7	2.1
2	1.3	1.9	2.2	11	1.9	2.3	2.4
3	1.7	1.4	2.6	12	0.5	1.6	1.2
4	0.8	2.1	1.9	13	2.2	2.6	2.9
5	2.0	2.6	2.4	14	2.0	2.4	2.3
6	1.0	1.8	1.4	15	1.9	2.3	2.6
7	1.5	2.1	2.2	16	2.1	2.3	2.7
8	1.4	2.4	2.0	17	1.6	2.3	2.3
9	1.5	2.1	2.1				

Again, as in the correlational procedure, statements 6 and 12 were identified as relatively poor with respect to discriminating power. There was no question, then, that statements 6 and 12 should not be retained in the scale. As tables 6-3 and 6-4 show, the results of the criterion of internal consistency were similar to that of the correlational procedure (rf. Edwards, 1957, p. 155; and Likert, 1932, p. 50). For example, in the secondary group, statement 3 was relatively poor, as was statement 12, although statement 6 appeared satisfactory. In the elementary group, statements 4, 6, and 12 were relatively low, although statement 17 appeared satisfactory. The only real discrepancy, then, between the correlational method and the criterion of internal consistency, was found in statement 17. According to the criterion of internal consistency, this statement discriminated well, across all groups.

Conclusions

The evidence in total, therefore, from a logical and empirical perspective, identified statements 3, 4, 6, and 12 as problematic, lacking acceptable discriminating power for all groups, and so were eliminated from the proposed scale. Statement 17, although the criterion of internal consistency identified it as having acceptable levels of discriminating power for all groups, was also rejected because the correlational method was identified by this study as the primary source of empirical evidence upon which to make decisions, and the correlation coefficient for the elementary group was unacceptably low. In addition, the results of the item analysis demonstrated that the numerical values were properly assigned to the responses, as there were no negative coefficients or negative critical ratios; and that, apart from statements 3, 4, 6, 12, and 17, the scale appeared to be appropriate for the intended grades. Although some variation existed between groups, the variation was slight and the correlation coefficients were at acceptable levels. Even in well developed scales, empirical evidence varies from one sample to another, (Anderson, 1981, pp. 139-141), and just because items discriminate differently for different age groups, this is not evidence of a subscale (Gardner, 1975, p. 104).

Finally, the statements appeared to be related to each other, the underlying assumption being that the statements meeting the tests are measures of a common variable (rf. Rundquist & Sletto, 1975, p. 306). There is a danger in this

assumption, however, because neither test guarantees that the statements are measuring one variable only. The reader must recognize that even though all final statements correlated satisfactorily with the total score, the possibility of subscales within the larger scale exists (Lemon, 1973, p. 181; Scott, 1968, p. 219). Statistical procedures are available to test for the evidence of subscales, such as factor analysis and Scott's formula for homogeneity; however, Scott suggests that these may not be necessary unless there is reason to suspect that the statements are not measuring a single attribute; e.g., attitudes toward the police. Gardner (1975) also states that although traditional methods of item analysis used for summated ratings scales cannot identify subscales, internal consistency is a reflection of a unidimensional construct (p. 106; rf. also Cronbach, 1951, pp. 320, 331; Guilford & Frutcher, 1973, p. 407; and Kaplan & Saccuzzo, 1982, p. 103).

Unidimensionality itself is perplexing, because theoretically one could classify and subclassify any abstract attribute indefinitely (Likert, 1932, p. 7). To conclude with absolute certainty that a scale is unidimensional, then, is difficult, especially since the attribute to be measured is not directly observable but must be inferred from certain kinds of behavior. According to Scott (1968), a scale is strictly unidimensional in the sense that "all items measure the intended attribute only"; more realistic, though, is that the set of items "reflects the intended trait more than it reflects any other trait"; however, "usually one is satisfied with a more

modest achievement" where, for most subjects, the total score represents, for the most part, the intended trait (p. 250).

In conclusion, emphasis in item analysis was placed on clearly reasoned theory, and secondly, on statistical analysis (rf. Anderson, 1981; and Gardner, 1975). The procedures used in the statistical analysis were a simple correlational procedure and the criterion of internal consistency. More sophisticated measures exist, such as factor analysis; however, there are certain advantages to simplicity, especially if it yields essentially the same results. For example, Borg and Gall (1983) suggest that "the student who processes his own research data on a desk calculator has a much better insight into what his data indicates..." (p. 853; rf. also Wideen, 1987). However, with respect to this study, the danger was that the procedures used did not guarantee the absence of subscales; notwithstanding, there was no reason to suspect that the scale was not unidimensional; i.e., for most students, the total score represented, for the most part, the intended trait.

Therefore, based on the findings of item analysis, 12 statements were selected as appropriate for the finished scale; however, 5 statements were negative and 7 statements were positive with respect to the psychological object of interest. Because of the criterion that a Likert-type scale consist of an approximately equal number of positively and negatively worded statements in order to ameliorate the acquiescence response set, it was necessary to eliminate one positively worded statement. Statement two was therefore eliminated, one reason being that it

had one of the lower coefficients of the positively worded statements, and a second reason being that it had a relatively low critical ratio for the elementary group. As a result, 11 statements were selected for the finished scale (rf. Appendix E).

CHAPTER VII

Reliability and Validity

Following item analysis, the construction of the proposed scale was complete. However, it was necessary to evaluate the scale, based on assessments of reliability and validity, in order to conclude on its potential usefulness. Because reliability is a logical prerequisite to validity, and because validity is difficult to establish in attitude scales, reliability was a primary consideration in this evaluation. As a result, this chapter first considers the issue of reliability, and secondly considers the issue of validity.

Reliability

Conceptual Definitions. Classical test score theory assumes that each person has a true score that could be identified except for errors in the measurement process. The difference between the "true score" and the observed score results from measurement error (Kaplan & Saccuzzo, 1982, p. 87). Notwithstanding that systematic errors are acknowledged in classical test theory, a major assumption is that errors of measurement are random. The implication is that errors will be normally distributed, allowing the assumption that the mean score of a measure will be the estimated true score and the remaining distribution of scores the result of random errors. Theoretically, then, the standard deviation of the distribution reveals the magnitude of measurement error, although repeated

administrations would be necessary to obtain this information. Another assumption, however, necessitated by the fact that this theory accommodates only one subject, is that the distribution of random errors will be the same for all subjects (Kaplan & Saccuzzo, p. 89). Succinctly defined, then, classical test theory uses the standard deviation of errors as the basic measurement of error, called "standard error of measurement." The problem with all of this, of course, is that to compute an estimated true score presents enormous practical problems; in practice, reliability coefficients are usually estimated from a single administration. The same principle holds true for the domain sampling model, where reliability coefficients are usually estimated from a single measure (rf. Kaplan & Saccuzzo, pp. 95-96).

Although most reliability coefficients, in reality, are correlations, it is helpful to conceptualize reliability in terms of a ratio. According to Kaplan and Saccuzzo (1982), "the reliability coefficient is the ratio of the variance of the true score of a test to the variance of the observed score," written as $R = \sigma_{T^2} / \sigma_{X^2}$ (p. 90; rf. also Borg & Gall, 1983, p. 283; Cronbach, 1970, p. 154; and Lemon, 1973, pp. 45-46). As the variance of the observed score approaches that of the true score, reliability concomitantly increases. This ratio may be thought of as a percentage (Kaplan & Saccuzzo, p. 90). For example, if the reliability coefficient was .8, 20% of the variation between the scores of the subjects would be assumed random, and 80% of the variation would be assumed true variation

between the subjects.

Operational Definitions. Before estimating reliability, according to Kaplan and Saccuzzo (1982), the source of measurement error (i.e., stability, consistency, equivalence) should first be identified (p. 114). Traditionally, researchers refer to three types of reliability: (1) stability, generally defined as the extent to which a measure is consistent over time; (2) consistency, generally defined as the extent to which responses to items of a measure are consistent with each other (internal consistency); and (3) equivalence, generally defined as the extent to which responses made to alternate forms of a measure are consistent with each other (rf. Anderson, 1981; Cronbach, 1951; Gronlund, 1985; and Lemon, 1973). Although these definitions imply differences in reliability, the real difference lies in emphasis and method (Lemon, p. 46).

Although measures of internal consistency are the most commonly used (Anderson, 1981, p. 108), it is important to be aware of the advantages and disadvantages of each method as they relate to attitude scales. Stability is estimated by the test-retest technique, where a scale is administered and readministered at a later date, the scores subsequently correlated to yield an estimate of reliability. Criticisms of this method are numerous. For example, it is inconvenient and at times impossible, because of the difficulties in arranging for two similar administrations for the same group (Cronbach, 1970, p. 298; Lemon, 1973, p. 47). In addition, there are

methodological problems, such as a subject's memory of the first administration confounding the results (Adams, 1982, p. 186; Kaplan & Saccuzzo, 1982, pp. 92-93; Lemon, 1973; p. 47; Scott, 1968, p. 256); and because of the fact that only the scores are correlated, it is unknown whether or not the observations which make up the scale are correlated (Lemon, p. 48). The only advantage to this technique is that stability over time is considered.

Internal consistency is usually estimated by the split-half technique, corrected by the Spearman-Brown formula, where reliability can be estimated from a single administration. According to Cronbach (1951), the split-half technique resolved most of the problems associated with the test-retest technique; however, a different "quality" of reliability is being measured; i.e., internal consistency (p. 298; rf. also Borg & Gall, 1983; and Likert, 1932). Although this method is recommended (rf. Kaplan & Saccuzzo, 1982, p. 97; Lemon, 1973, pp. 49-50; and Scott, 1968, p. 257), it is criticized because it can yield different reliability coefficients, depending on how the scale is divided (Cronbach, p. 300; Lemon, pp. 49-50). Additionally, because the split-half technique requires that the items be divided equally, if a scale consists of an odd number of items, the deletion of an item in order to compute the reliability coefficient is necessary (Likert, pp. 30-31).

Because of the criticisms of the split-half technique, Cronbach's coefficient alpha (Cronbach, 1951) is often recommended, although researchers such as Kaplan and Saccuzzo

(1982) believe that the alpha technique is only advantageous if the two halves of the scale have unequal variances (p. 98). The coefficient alpha "tells how well scores obtained by testing under just one condition ... represent universe scores" (Cronbach, p. 160). In other words, this reliability coefficient "tells what proportion of the observed variance score is non-error variance" (Cronbach, p. 165), which is the theory of reliability in a nutshell. The advantage of the alpha coefficient is that it gives the average of all possible split-half coefficients for a given test, and it is "a good index of homogeneity" (Cronbach, 1970, pp. 300, 331). Another advantage is that if the scale consists of an odd number of items, one item does not have to be deleted because the longer version of the coefficient alpha formula ($R = (N / N - 1) (S^2 - \sum S_i^2 / S^2)$) (rf. Kaplan & Saccuzzo, p. 103) incorporates all responses to all items. Finally, it is recommended by most researchers for use with attitude scales (rf. Adams, 1982; Anderson, 1981; Kaplan & Saccuzzo; Lemon, 1973; and Scott, 1968).

Equivalence is estimated by administering alternate forms of the scale, either at the same time or at different times, the reliability coefficient estimated by correlating the scores between the two forms. If the two forms are generated by techniques such as the split-half technique, there is no methodological problem in correlating total scores, as there is in the test-retest technique. According to Kaplan and Saccuzzo (1982), Anderson (1981), Lemon (1973), and Gronlund (1985), this

technique is the most rigorous and the best as it takes into account both stability and internal consistency; however, it has practical limitations which are similar to the test-retest method, including the necessity of developing alternate forms (Kaplan & Saccuzzo, p. 94; Lemon, p. 151).

Estimates of Scale Reliability. With these qualifications in mind, this study depended on measures of internal consistency in order to estimate reliability.

Table 7-1
Statistical Description of the Sample

Grades	Mn Age	N (Male)	N (Fem)	N (Total)
4	9.4	15	13	28
5	10.5	17	11	28
6	11.6	19	11	30
7	12.4	12	18	30
8	13.6	13	16	29
9	14.5	16	8	24
10	15.2	16	11	27
11	16.4	9	15	24
12	17.2	12	14	26

Table 7-2
Statistical Summaries of the Sample

Grades	Mn Age	N (Male)	N (Fem)	N (Total)
El:4-7	11.0	63	53	116
Sec:8-12	15.3	66	64	130
Tot:4-12	13.3	129	117	246

What was of concern was that the responses to the proposed scale (rf. Appendix E) were consistent and therefore repeatable

(Moore, 1979, p. 117). The data in Tables 7-1 and 7-2 show that the sample was representative of the group of interest; however, as in item analysis, sample size restricted an analysis to the group as a whole and to the elementary and secondary groups, as previously discussed in the chapter on methodology.

Because the total sample exceeded the minimum requirements as defined in the literature and was relatively large, it was used as the primary source of data on which to base the reliability coefficients. In addition, because the sample size of the elementary and secondary groups met or exceeded the minimum requirements, they also were analyzed to determine if their results were generally consistent with the total sample, and so yielded additional data upon which to make decisions.

Table 7-3
Reliability Coefficients

	Grades		
Technique	El (4-7)	Sec (8-12)	Total (4-12)
Split-half	.78	.82	.84
Split-half*	.88	.90	.91
Alpha	.82	.82	.84

*Corrected by the Spearman-Brown formula (R = 2r / 1 + r).

Table 7-3 shows the estimated reliability coefficients for the elementary, secondary, and total groups. The split-half coefficient was computed by using the odd-even method (rf. Kaplan & Saccuzzo, 1982, p. 97), deleting the last item (rf. Likert, 1932, pp. 30-31); and the alpha coefficient was computed

by using the long formula vis-a-vis the split-half formula (rf. Kaplan & Saccuzzo, pp. 99, 103), so that all items were included.

According to Borg and Gall (1983), the range of reliability coefficients for published attitudes scales is from a low of .47 to a median of .79 to a high of .98 (p. 282). The necessary level of reliability is determined by the proposed use for the scale. For example, if the scale is for research purposes, a coefficient between .7 and .8 is satisfactory (rf. Borg & Gall, p. 292; and Kaplan & Saccuzzo, 1982, p. 106). Unless the scale is to be used to make decisions about individuals, to attempt to achieve a coefficient of .9 is usually not worth the time and money (Kaplan & Saccuzzo, p. 106; rf. also Cronbach, 1983, p. 103). Anderson (1981), supplying more detail, states that acceptable levels of reliability varies according to type. For example, for measures of internal consistency, .8 is satisfactory; for measures of equivalence, .7 is satisfactory; and for measures of stability, .7 is also satisfactory (Anderson, p. 110).

Another point to be noted is that reliability is affected by a number of factors. For example, reliability is related to the length of a test. This is because the increased number of items results in a better sampling of the domain, the result being a more reliable estimate of the "true score" (Borg & Gall, 1983, p. 292; Cronbach, 1951, p. 323, & 1970, p. 166; Kaplan & Saccuzzo, 1982, pp. 95, 106). In addition, reliability level is also a function of the variability (spread between scores)

between subjects (Cronbach, 1970, p. 165; Scott, 1968, p. 256), and according to Anderson (1981), the reliability coefficient can vary by as much as .4 or more from sample to sample, even in well developed scales (p. 140). Finally, as previously implied, the type of measure by which reliability is estimated is related to the size of the reliability coefficient, where methods of equivalence with a time interval typically provide the lowest level of reliability (Gronlund, 1985, p. 105).

Although, all things being equal, more items yield higher levels of reliability, it is more precise to say that reliability is increased by the number of response opportunities (Anderson, 1981, p. 251; Cronbach, 1970, p. 165). With respect to the summated ratings scale, the five-point response format has been found to yield the best levels of reliability (Adams, 1982, p. 186; Anderson, p. 251; Likert, 1932). However, notwithstanding that reliability can be increased by increasing the opportunities for responses, Likert-type attitude scales with acceptable levels of reliability can be constructed with as few as 5 items (Edwards, 1957), although a minimum of 10 items is often recommended (Borg & Gall, 1983, p. 423; Gronlund, 1985, p. 419).

A case in point apropos to this study is that of Phillips and Coates (1971), who constructed an 11 item standard summated ratings scale designed to assess attitudes toward the police, obtaining a reliability coefficient of .81 (split-half technique). Regardless of the theoretical advantages of a lengthy test, paradoxically, a short test is often recommended

because the best items are retained (Borg & Gall, p. 292), resulting in a more homogeneous scale; and if acceptable standards have been met, according to Cronbach (1951), increasing the test length often serves no useful purpose (p. 323). Finally, as has been previously noted in Chapter VI, relatively short scales are recommended for children (Anderson, 1981, p. 69).

Validity

Only after a demonstration of reliability can validity be seriously considered, because a scale that is unreliable cannot provide valid information about the construct being measured (Gronlund, 1985, p. 87). However, reliability does not guarantee validity, as the scale might not be measuring the construct for which it was intended, notwithstanding consistent results.

Conceptual Definition. Simply stated, a measure is valid to the extent that it does what it is designed to do (Anderson, 1981, p. 90; Lemon, 1973, p. 37). For example, a thermometer is a valid instrument to use for assessing temperature. Validity, however, cannot be verified without prior information. In the example of the thermometer, information must be available about temperature; and in the case of attitude scales, information must be available about attitudes. Because attitudes cannot be directly observed but must be inferred by some defined behavior, and because no consensus exists as to the construct of attitude,

a demonstration of validity is difficult.

Operational Definitions. The general trend, then, in an effort to establish the validity of an attitude scale, has been "to establish strong positive correlations between measures in order to show that they measure the same attitude" (Lemon, 1973, p. 53). According to Lemon, the methods of predictive and construct validity are based on this principle, a strategy that has been labelled convergent validation (p. 53). This, then, raises the issue of types of validity, the differences, as in reliability, being one of emphasis. Here, Anderson (1981) makes an important distinction between the types of validity: logical validity, which includes face and content validity; and empirical validity, which includes predictive and construct validity (logically including convergent validity) (p. 106).

Assessment of Scale Validity. Face validity is the extent to which the instrument appears relevant to an interested observer (Cronbach, 1970, p. 183; Lemon, 1973, p. 37). With respect to the scale proposed in this study, this criterion was met. This scale was read and reviewed by literally hundreds of people, including students in grades 4 to 12, university students, teachers, university professors, and police officers, and its face validity was never doubted. Content validity is not frequently encountered in attitude assessment, but is defined as the extent to which the items constituting the proposed scale adequately represent the domain of interest

(Lemon, pp. 39-40). Again, this criterion was met. The psychological object of interest for this study was the attitudes of students toward the police, the theoretical basis for selecting items that would adequately represent this domain being the principles of Sir Robert Peel, as discussed in chapter IV. The only principle not apparently addressed was number nine: to recognize that the test of police efficiency is the absence of crime and disorder. This principle appeared to demand a statement of fact rather than attitude; however, statement number 10 in the proposed scale, "The police are doing a good job," indirectly addressed this principle.

With respect to logical validity, then, the proposed study clearly defined the psychological object of interest, the target, and the theory, carefully outlined the steps used in constructing the scale, and logically justified statement selection. According to Anderson (1981), these issues must be addressed in any attempt to claim logical validity, issues to which this study carefully attended.

Predictive validity, according to Lemon (1973), "is the degree to which it enables the investigator to predict the value of some criterion" (p. 32; rf. also Moore, 1979). For example, attitudes toward religion might be correlated to church attendance; or, attitudes toward some ideology might be correlated to group membership. Because both predictive validity and concurrent validity use some defined criterion as a basis to judge validity, to distinguish between the two is somewhat unnecessary (Lemon, p. 39). With respect to attitudes

toward the police, Phillips and Coates (1971), to demonstrate predictive validity for their scales, identified groups (which included police officers and undergraduate black students) and predicted that their scale would discriminate between them in a manner consistent with previous research. Because the scale discriminated as predicted, Phillips and Coates concluded that "criterion group tests support the validity of the scales" (p. 8).

Table 7-4
Data Summaries

	Statistics		
Grades	Mn	SD	Var
El:4-7	45.3	7.5	56.0
Sec:8-12	38.3	7.7	58.6
Tot:4-12	41.6	8.3	69.5

Similarly, this study, because data (rf. Table 7-4) were available on the attitudes of students in grades 4 to 12 as a result of the reliability research, predicted that the proposed scale would accurately discriminate between the attitudes of the elementary group and the secondary group of students. With respect to the independent variable, attitudes toward the police, and the dependent variable, grade, there is general consensus in the literature that the attitudes of secondary students are more negative than the attitudes of elementary students, and this was the expectation of this study.

As can be seen in Table 7-4, the attitudes of the

secondary group were apparently more negative than the attitudes of the elementary group (scores computed by standard procedures for summated ratings scales).

Table 7-5
t-test Comparing Independent Means

Comparison	t-test	Critical Value
Between El and Sec	7.2*	3.29

*p < .0005

A t-test (rf. Table 7-5) comparing independent means showed this difference to be statistically significant, p (t_{244} = 7.2) < .0005, supporting the predictive validity of the proposed scale.

Construct validity is more complex than the other forms of validity, and is defined by Lemon (1973) as "an index of the degree to which an individual's performance on ... [a] measure can be ascribed to this construct" (p. 41). With respect to the construct of attitude, it is an abstraction, necessarily defined in terms of properties for the purposes of measurement. Therefore, a demonstration of construct validity is necessarily dependent on the researcher's conceptualization of attitude, about which little agreement exists, and on the researcher's concomitant operational definition.

Validity, according to Moore (1979), "is a simple but slippery idea," especially in the social sciences, where the property to be measured "is so fuzzy that reasonable persons can

disagree on the validity of a variable as a measure of that property" (p. 114). As a result, construct validity, he goes on to say, is the attempt "to show a connection between the measuring variable and other variables that ought to be connected with the original fuzzy idea" (p. 116). This study attempted to define "the original fuzzy idea," but offers no evidence of construct validity, apart from a somewhat superficial demonstration of predictive validity.

Conclusions.

The primary consideration of this chapter was scale reliability; and based on the literature reviewed, the estimated reliability coefficients are adequate for the purpose of the proposed scale. In addition, based on the apparent consistency between the empirical data of the elementary and secondary groups, and the apparent consistency between the subgroups and the total group, the proposed scale (rf. Appendix E) is appropriate for the range of grades for which it is intended (rf. Anderson, 1981, p. 122).

With respect to validity, although a case for logical validity was made, the case for empirical validity was limited to a simple demonstration of predictive validity. As a result, definitive conclusions on validity cannot be made. Because of the contradiction and controversy involved in the whole area of attitude, it is not uncommon for researcher's to concentrate on reliability (Scott, 1968, p. 256). For example, Portune (1965), commenting on the validity of his attitude scale,

quoting Horton, Jr. (1963), said, "Having defined belief as the subject of research ... the question of validity might be considered as being resolved into the question of reliability alone" (p. 46). Although this study makes no such conclusion, the issue of validity is left for future research.

CHAPTER VIII

Summary and Conclusions

Because of the role that the police play in society, one of the primary concerns of a democratic society should be the existing relationship between the police and the public, including its young people, because attitudes are formed early and may persist, justified or not, throughout life. Therefore, schools, as public institutions designed to serve society, have an important responsibility because of their enormous potential to influence behavior and attitudes. However, serious deficiencies, such as inadequate assessment techniques and the ignoring of the attitudes of elementary students, exist in the subject field of attitudes toward the police. In an effort to address these deficiencies, recognizing that research is only as good as the methodologies and assessment techniques used, the purpose of this study was to construct an attitude scale (summated ratings type) that would accurately assess the attitudes of students in grades 4 to 12 toward the police.

The procedures used to construct the attitude scale were both quantitative and qualitative in their approach. Quantitative methodology with respect to instrument construction is well known. A pool of items is selected, item analysis is conducted by means of statistical procedures, followed by an estimate of reliability, and a consideration of validity. However, well developed foundational theory is often lacking, which according to Gardner (1975) and White and Menke (1978)

accounts for a serious flaw in current attitude assessment scales, resulting in research in the field of criminal justice not being taken seriously. According to Gardner, "effective research in this field [attitude measurement] demands the construction of attitude scales which clearly reflect some underlying theoretical construct" (p. 101).

In this study, careful attention was paid to theory, the proposed scale finding its theoretical basis in the principles of Sir Robert Peel. These principles were critically important, influencing the construction of the scale from start to finish. For example, after an item pool was assembled, each statement was examined with respect to its consistency to the defined theoretical foundation of the study, and only those statements which were consistent with this foundation were retained. In addition, in item analysis, careful attention was paid to conducting a logical analysis, often omitted in instrument development. Here, not only was each statement subjected to scrutiny as to whether or not it met the criteria for inclusion in a Likert-type scale, but also as to whether or not it met the theoretical criteria.

With respect to methodology, then, this study followed the principle of theory first, confirmation by statistics second, a principle which Gardner (1975) states has clearly been "ignored" (p. 102). As a result, measurement in this study reflected clear concepts (rf. Gardner, p. 102); in other words, an attempt was made to define exactly what variable the scale measured. An important distinction, therefore, was made between the

theoretical foundation of the scale and attitude scale theory. The theoretical foundation must logically precede the application of scale theory, as "the absence of any theoretical rationale makes attitude scale theory entirely inapplicable, and statistical procedures such as summing item scores, split-half reliability, and so on ... completely irrelevant" (Gardner, p. 102).

Considering the importance that theory played in the construction of the proposed scale, and that out of 19 opinion statements that appeared to reflect adequately the domain of interest only 11 were retained in the final form of the scale, and that these 11 were refined through rewriting in order to establish readability at grade four, it is necessary to re-examine the finished scale in order to determine if it reflects the domain of interest as originally intended. For example, according to Peel's principles, the first responsibility of the police in a democratic society is to prevent crime and disorder, a responsibility that is of great concern today, and which is often stated in terms of whether or not the police are doing a good job. This, then, is reflected in statement number 10, "The police are doing a good job" (rf. Appendix E).

The second responsibility of the police, according to Peel, is to recognize that policing authority in a democracy is dependent upon public approval and cooperation; i.e., that the strength of the police is not physical, but moral. This principle, probably not overtly recognized by most people, is

implicit in the proposed scale. The purpose of the scale is to facilitate the assessment of the existing relationship between the police and the public, and if the issues reflected in the statements, such as police thoughtfulness for people, police friendliness to people, police treatment of people, and police impartiality with people, are not positively perceived by the public, it is quite likely that this moral strength, if not lost, will be seriously impaired.

Peel's principles number three, to secure and maintain the respect and approval of the public, number four, to recognize that public cooperation is directly related to the amount of physical force needed to achieve social order, and number five, to seek and preserve public favor by being sensitive to public opinion, by being courteous and friendly, and by being impartial, are closely related to principle number two. For example, the need to be sensitive to public opinion is addressed by the purpose of the scale, and concerns for courtesy and friendliness are reflected in statements number two, "The police are always mad at kids," number four, "The police are friendly," and number six, "The police are pretty nice people." Police thoughtfulness is reflected in statements number five, "The police care for young people," and nine, "The police don't care what happens to you after they pick you up," and police impartiality is reflected in statement number one, "Young people are treated fairly by the police."

Peel's principles number six, to use a minimum of force, and only when necessary, and number eight, to refrain from

usurping the power of the judiciary, is clearly reflected in statement number 11, "The police enjoy kicking people around." These two principles, within a larger perspective, also deal with the general issue, abuse of authority, that is so much of a concern in a democratic society, and is also indirectly reflected in statements such as those dealing with fairness (number one), friendliness (number four), thoughtfulness (numbers five and nine), and impartiality (number one).

The responsibility of the police to recognize that the test of police efficiency is the absence of crime and disorder, found in principle number nine, is not directly found in the scale, although such a statement could easily be written. The problem with such a statement, though, is that it would appear to be a statement of fact rather than opinion, a conclusion supported by the fact that the statements "Without the police, there would be crime everywhere," and "The police are not needed in our country" failed to discriminate between persons of differing opinions (rf. Chapter VI, Item Analysis). However, statement number 10, "The police are doing a good job" indirectly addresses the issue of police efficiency.

Finally, Peel's principle number seven, "To maintain at all times a relationship with the public that gives reality to the historic tradition that the police are the public and that the public are the police; the police being only members of the public who are paid to give full-time attention to duties which are incumbent, on every citizen, in the interest of community welfare and existence" (rf. Radelet, 1973, pp. 4-5), along with

principle number two, is the core of policing in a democracy. As in principle number two, where most people probably do not overtly recognize that the strength of the police in a democracy is moral, most people probably do not overtly identify with the police. However, as in principle number two, this principle is addressed in the purpose of the scale itself. In a democracy, public approval, support, and willing cooperation would not be granted to the police if they were "dumb," (rf. statement number three), generally a pejorative designation for the dull witted, if they were dishonest (rf. statement number seven), or if they were "just as bad as the people they put in jail" (rf. statement number eight).

In conclusion, from a logical perspective, the opinion statements constituting the proposed scale appear to reflect adequately the domain of interest, the finished product not substantially different from that originally intended. From an empirical perspective, the statements appear interrelated and homogeneous, demonstrated by the correlations of the selected items to the battery score, and by the reliability coefficients, which are measures of internal consistency (rf. Cronbach, 1951, pp. 320, 331; Gardner, 1975, p. 106; Guilford & Frutcher, 1973, p. 407; and Kaplan & Saccuzzo, 1982, p. 103).

In closing, this study defined its objectives, which included a description of the population of interest, the trait to be measured, the technique selected for the measurement process, and the scoring process. In addition, this study addressed the issues of readability and item analysis; and is

now satisfied to conclude with a scale, built on well developed theory, that has a satisfactory level of reliability. With respect to validity, although this study demonstrated logical validity and a superficial level of empirical validity, a convincing demonstration is left to subsequent research. In the field of criminal justice, however, such a demonstration will be difficult because current research is marked by contradiction and controversy, as, indeed, is the case in the field of attitude measurement itself. The challenge for future research, then, is clear, as a resolution of the fundamental, contentious issues would permit getting on with what should be one of the primary concerns of a democratic society--the existing relationship between the police and the public.

APPENDIX A

Student Interview Guide

Objectives

1. To confirm readability level.
2. To determine if elementary students can discriminate on the five-point Likert scale.
3. To facilitate the selection of statements, based on readability.

Directions

The interviewer follows the Student Interview Guide, asking questions according to the format outlined, and scoring the appropriate responses in the spaces provided.

Part I: Demographics

Date:_____ Start time:_____ Finish time:_____

School: _____

Principal: _____ Teacher: _____

Location of Interview: _____

Interviewer: _____ Race: _____

Age: _____ Sex: _____

Respondent: _____ Race: _____

Age (year & month): _____ Grade: ___ Sex: ___

Teacher's/Principal's general description of respondent (Re: intelligence, academic performance, representativeness of peer group):

_____

_____

_____

_____

Part 2: Introduction

1. Hi (state first name of student). My name is _____. How are you today? I guess your teacher told you that I am a police officer.

   Engage in some friendly conversation for a couple of minutes, putting the student at ease, creating a friendly, non-threatening atmosphere.

2. The reason that I asked to talk to you today is that I need your help. I am going to school, just like you are, and I am trying to make up a test that will help people to understand how students feel or think about the police.

   Does that sound interesting?

   Now don't worry; you do not have to tell me how you feel about the police. But you can really help me by reading the sentences on the test, by telling me what you think the sentences mean, and if you think that other students in your grade would be able to understand the sentences.

   OK? And don't worry if you make a mistake; it's OK to make mistakes, and I won't be telling anyone.

   Also, I would like you to play a little game with me, where you would pretend to be an actor, like on TV, and tell me how different people might answer some of the questions. That will be fun. OK?

   Let's begin then.

Part 3: Readability of Statements

Directions
Give the student the sheets of paper on which are typed the 19 statements and responses; the students are not to complete the responses, except when role playing (Part 4).

For the remainder of the interview, the following symbols apply:
S Make a statement
Q Ask a question
R Record the response

****

S Now, here, I am going to ask you to read the sentences aloud, one at a time. After you read each sentence, I am going to ask you what you think the sentences mean, and if you think that other students in your grade would be able to understand the sentences. OK?

1. S Please read sentence number 1 aloud.
   R Mistakes:
   ( ) Yes.
   If yes, which words? _____
   ( ) No.
   Q Were there any words in the sentence that you did not understand?
   R ( ) Yes.
   If yes, which words? _____
   ( ) No.
   Q What do you think the sentence means?
   R Ability to restate the statement in a manner consistent with its original meaning.
   ( ) Yes.
   ( ) No.
   If no, how was the statement restated? _____

   _____
   Q Do you think that most students in your grade would be able to understand this sentence?
   R ( ) Yes.
   ( ) No.
   ( ) Maybe.
   S OK. Let's go to sentence number 2

2. S Please read sentence number 2 aloud.
   R Mistakes:
   ( ) Yes.
   If yes, which words? _____

( ) No.

Q Were there any words in the sentence that you did not understand?

R ( ) Yes.
If yes, which words? _____
( ) No.

Q What do you think the sentence means?

R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

_____

Q Do you think that most students in your grade would be able to understand this sentence?

R ( ) Yes.
( ) No.
( ) Maybe.

S OK. Let's go to sentence number 3

3. S Please read sentence number 3 aloud.

R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.

Q Were there any words in the sentence that you did not understand?

R ( ) Yes.
If yes, which words? _____
( ) No.

Q What do you think the sentence means?

R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

_____

Q Do you think that most students in your grade would be able to understand this sentence?

R ( ) Yes.
( ) No.
( ) Maybe.

S OK. Let's go to sentence number 4

4. S Please read sentence number 4 aloud.

R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.

Q Were there any words in the sentence that you did not understand?

R ( ) Yes.

If yes, which words? _____
( ) No.

Q What do you think the sentence means?

R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

_____

Q Do you think that most students in your grade would be able to understand this sentence?

R ( ) Yes.
( ) No.
( ) Maybe.

S OK. Let's go to sentence number 5

5. S Please read sentence number 5 aloud.

R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.

Q Were there any words in the sentence that you did not understand?

R ( ) Yes.
If yes, which words? _____
( ) No.

Q What do you think the sentence means?

R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

_____

Q Do you think that most students in your grade would be able to understand this sentence?

R ( ) Yes.
( ) No.
( ) Maybe.

S OK. Let's go to sentence number 6

6. S Please read sentence number 6 aloud.

R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.

Q Were there any words in the sentence that you did not understand?

R ( ) Yes.
If yes, which words? _____
( ) No.

Q What do you think the sentence means?

R Ability to restate the statement in a manner consistent

with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

_____

Q Do you think that most students in your grade would be able to understand this sentence?
R ( ) Yes.
( ) No.
( ) Maybe.
S OK. Let's go to sentence number 7

7. S Please read sentence number 7 aloud.
R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.
Q Were there any words in the sentence that you did not understand?
R ( ) Yes.
If yes, which words? _____
( ) No.
Q What do you think the sentence means?
R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

_____

Q Do you think that most students in your grade would be able to understand this sentence?
R ( ) Yes.
( ) No.
( ) Maybe.
S OK. Let's go to sentence number 8

8. S Please read sentence number 8 aloud.
R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.
Q Were there any words in the sentence that you did not understand?
R ( ) Yes.
If yes, which words? _____
( ) No.
Q What do you think the sentence means?
R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

Q Do you think that most students in your grade would be able to understand this sentence?

R ( ) Yes.
( ) No.
( ) Maybe.

S OK. Let's go to sentence number 9

9. S Please read sentence number 9 aloud.

R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.

Q Were there any words in the sentence that you did not understand?

R ( ) Yes.
If yes, which words? _____
( ) No.

Q What do you think the sentence means?

R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

Q Do you think that most students in your grade would be able to understand this sentence?

R ( ) Yes.
( ) No.
( ) Maybe.

S OK. Let's go to sentence number 10

10. S Please read sentence number 10 aloud.

R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.

Q Were there any words in the sentence that you did not understand?

R ( ) Yes.
If yes, which words? _____
( ) No.

Q What do you think the sentence means?

R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

Q Do you think that most students in your grade would be able to understand this sentence?

R ( ) Yes.
( ) No.
( ) Maybe.
S OK. Let's go to sentence number 11

11. S Please read sentence number 11 aloud.
R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.
Q Were there any words in the sentence that you did not understand?
R ( ) Yes.
If yes, which words? _____
( ) No.
Q What do you think the sentence means?
R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

_____

Q Do you think that most students in your grade would be able to understand this sentence?
R ( ) Yes.
( ) No.
( ) Maybe.
S OK. Let's go to sentence number 12

12. S Please read sentence number 12 aloud.
R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.
Q Were there any words in the sentence that you did not understand?
R ( ) Yes.
If yes, which words? _____
( ) No.
Q What do you think the sentence means?
R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

_____

Q Do you think that most students in your grade would be able to understand this sentence?
R ( ) Yes.
( ) No.
( ) Maybe.
S OK. Let's go to sentence number 13

13. S Please read sentence number 13 aloud.
R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.
Q Were there any words in the sentence that you did not understand?
R ( ) Yes.
If yes, which words? _____
( ) No.
Q What do you think the sentence means?
R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

_____

Q Do you think that most students in your grade would be able to understand this sentence?
R ( ) Yes.
( ) No.
( ) Maybe.
S OK. Let's go to sentence number 14

14. S Please read sentence number 14 aloud.
R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.
Q Were there any words in the sentence that you did not understand?
R ( ) Yes.
If yes, which words? _____
( ) No.
Q What do you think the sentence means?
R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

_____

Q Do you think that most students in your grade would be able to understand this sentence?
R ( ) Yes.
( ) No.
( ) Maybe.
S OK. Let's go to sentence number 15

15. S Please read sentence number 15 aloud.
R Mistakes:
( ) Yes.

If yes, which words? _____
( ) No.

Q Were there any words in the sentence that you did not understand?

R ( ) Yes.
If yes, which words? _____
( ) No.

Q What do you think the sentence means?

R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

_____

Q Do you think that most students in your grade would be able to understand this sentence?

R ( ) Yes.
( ) No.
( ) Maybe.

S OK. Let's go to sentence number 16

16. S Please read sentence number 16 aloud.

R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.

Q Were there any words in the sentence that you did not understand?

R ( ) Yes.
If yes, which words? _____
( ) No.

Q What do you think the sentence means?

R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

_____

Q Do you think that most students in your grade would be able to understand this sentence?

R ( ) Yes.
( ) No.
( ) Maybe.

S OK. Let's go to sentence number 17

17. S Please read sentence number 17 aloud.

R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.

Q Were there any words in the sentence that you did not understand?

R ( ) Yes.
If yes, which words? _____
( ) No.
Q What do you think the sentence means?
R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

_____

Q Do you think that most students in your grade would be able to understand this sentence?
R ( ) Yes.
( ) No.
( ) Maybe.
S OK. Let's go to sentence number 18

18. S Please read sentence number 18 aloud.
R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.
Q Were there any words in the sentence that you did not understand?
R ( ) Yes.
If yes, which words? _____
( ) No.
Q What do you think the sentence means?
R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

_____

Q Do you think that most students in your grade would be able to understand this sentence?
R ( ) Yes.
( ) No.
( ) Maybe.
S OK. Let's go to sentence number 19

19. S Please read sentence number 19 aloud.
R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.
Q Were there any words in the sentence that you did not understand?
R ( ) Yes.
If yes, which words? _____
( ) No.
Q What do you think the sentence means?

R Ability to restate the statement in a manner consistent with its original meaning.
( ) Yes.
( ) No.
If no, how was the statement restated? _____

_____

Q Do you think that most students in your grade would be able to understand this sentence?
R ( ) Yes.
( ) No.
( ) Maybe.
S OK. Now let's look at the answers underneath each sentence.

Part 3: cont.: Readability of Responses

S Now do you see the words underneath each sentence?
(Pause)
Please read these words aloud.
R Mistakes:
( ) Yes.
If yes, which words? _____
( ) No.
Q Are there any words here that you do not understand?
R ( ) Yes.
If yes, which words? _____
( ) No.
Q Do you think that most students in your grade would be able to understand these words?
R ( ) Yes.
( ) No.
( ) Maybe.
Q What do you think that the word "unsure" means?
R Ability to restate, using at least one synonym (e.g., no idea, don't know, undecided, etc.).
( ) Yes.*
( ) No.**
If no, what meaning was given? _____

_____

* If yes, this section is concluded. Go on to Part 4.

** If no, proceed with the following questions.

Q Instead of using the word "unsure," would the word "undecided" be easier?
R ( ) Yes.
( ) No.
( ) The same.
Q What about using "don't know"? Would that be easier than the word "unsure"?
R ( ) Yes.
( ) No.
( ) The same.
Q Which word do you think would be the best one to use, so that students in your grade would be able to understand it?
R ( ) Unsure.
( ) Undecided.
( ) Don't know.

Part 4: Differentiation Between the Responses

S Now, here, I would like you to play a little game with me, where you will pretend to be an actor, like on TV or in the movies, and to tell me how different people might answer some of the questions. This will be fun.

Q OK?
R ( ) Yes.
( ) No.
If no, this Part is concluded. Go on to Part 5.

S Who is your favorite actor on TV? Now just like the people who act in TV shows, or in the movies, you need to to know what kind of person you're supposed to be. Well, you're going to need to be a really good actor, because you have five different persons to play:

1. One person, s/he (relate to student) really likes the police a whole lot.
2. The second one likes the police, but not as much as the first one.
3. The third person is unsure, s/he doesn't really know if s/he likes the police or not.
4. The fourth person does not like the police.
5. And the fifth person, well, s/he really hates the police.

Q OK?
R ( ) Yes.
( ) No.
If no, this Part is concluded. Go on to Part 5.

Q Do you have any questions? Do you understand?
R ( ) Yes.
If yes, attempt to clarify any misunderstandings.
( ) No.

1. S Now, pick any sentence you want.

R # _____
S PRETEND NOW that you really like the police a whole lot. You think that they are really wonderful people, the best in the whole world.
Now read the sentence to yourself.
Q Now tell me, how would such a person answer that sentence?
R ( ) SA; ( ) A; ( ) U; ( ) D; ( ) SD.
S Very good. You're doing a great job. Let's try another.

2. Now, pick any sentence you want.

R # _____
S PRETEND NOW that you really do not know if you like the police. It's not that you do not like them, it's just that you don't know how you feel about them.
Now read the sentence to yourself.
Q Now tell me, how would such a person answer that sentence?
R ( ) SA; ( ) A; ( ) U; ( ) D; ( ) SD.
S Very good. You're a great actor. Let's try another one.

3. S Now, pick any sentence you want.

R # _____
S PRETEND NOW that you like the police, not a whole lot, but you like them. You think that they are nice people.
Now read the sentence to yourself.
Q Now tell me, how would such a person answer that sentence?
R ( ) SA; ( ) A; ( ) U; ( ) D; ( ) SD.
S Very good. You're doing a great job. Let's try another.

4. S Now, pick any sentence you want.

R # _____
S PRETEND NOW that you really hate the police. You think that they are really awful people, the worst in the whole world.
Now read the sentence to yourself.
Q Now tell me, how would such a person answer that sentence?
R ( ) SA; ( ) A; ( ) U; ( ) D; ( ) SD.
S Very good. Now I have one more person for you to play.

5. S Now, pick any sentence you want.

R # _____
S PRETEND NOW that you do not like the police, not that you hate them, but you do not like them. You think that they are not nice people.
Now read the sentence to yourself.
Q Now tell me, how would such a person answer that sentence?
R ( ) SA; ( ) A; ( ) U; ( ) D; ( ) SD.
S Excellent. You are a great actor, and you played the parts really well.

S We're finished playing the game now, and so you don't have to pretend anymore? OK? That was fun.

Part 5: Post-interview Conversation

S I want to thank you very much for helping me today. You were a great help, and a very understanding person.

Q Were there any questions you wanted to ask me, now?

Q What class are you going to now?
* Engage here in a minute or so of informal conversation, talking about anything of interest to the student.

S Well, I guess your teacher will be waiting for you. Thanks again, (state first name of student). Bye.

APPENDIX B

Survey Manual for the Prototype

Survey Manual

How Students Feel About the Police

(A survey of student attitudes toward the police)

Introduction

My name is Paul Tinsley, and I am a graduate student at Simon Fraser University, Faculty of Education. I am also a police officer with the Matsqui Police Service. Presently, I am engaged in educational research in this school district, and I am especially interested in the attitudes of students toward policing as it exists in a democracy.

I would very much appreciate your help in administering this survey. No names will be used, and the results will be scored in group form only. The results of this study are extremely important, as I am trying to assess the validity of the survey. The importance of your cooperation as the administrator of the survey cannot be overstated.

Should you have any questions, or would like additional information regarding the study or the results, please feel free to contact me (phone the Matsqui Police at 859-5225, or my home at 859-0466).

Thankyou.

Directions for Administration

As this is a survey of attitudes, the biggest problem is to obtain honest answers, as the students might have a tendency to respond to the statements in a manner they they feel is socially desirable.

Research has shown that this problem can be overcome through attention to a few details in administration, so I encourage you to follow these directions carefully.

1. You (the teacher administering the survey) are the most important factor. A teacher has a strong rapport with his/her students, and the students will most often respond as instructed.

2. Ensure that conditions are conducive to the administration, and that enough time is set aside (12 minutes for secondary students; 15 minutes for elementary students).

3. Assure the students that the results are confidential. No names are requested, and it is the group scores that are important.

4. Assure the students that this is not a test, that there are no right or wrong answers. What is requested is their honest answers.

5. Convince the students of the importance of the research, and the importance of their cooperation.

6. Read the directions found on page 1 of the survey aloud. Do the example with the class, and answer any clarifying questions.

7. Direct the students to complete the survey on their own, responding once to each statement.

Class Profile

(To be completed by the class teacher)

School:_____ Teacher:_____

Grade:_____ N of Students:_____ Male:_____ Female:_____

Racial:_____

_____

Intelligence:_____

_____

Academic Performance:_____

_____

_____

Behavioral/Attitude:_____

_____

Description of Area:_____

_____

Family Background:_____

_____

Anything Unusual:_____

_____

_____

_____

Thankyou!

APPENDIX C

Prototype Scale

How Students Feel About the Police

* You are not asked to sign your name.

Age:_____ Grade:_____ Sex:_____

School:_____

Directions (To be read aloud to the class.)

This is not a test. There are no right or wrong answers. Just answer each sentence as honestly as you can.

Each sentence asks how you feel about the police. You are to tell how you feel by making an X by one of the answers found under each sentence.

Here is an example:

The police work hard.
( ) Strongly Agree
( ) Agree
( ) Unsure
( ) Disagree
( ) Strongly Disagree

Which answer tells best how you feel about the sentence? Put an X by only one answer.

Please read each sentence carefully. Remember, the best answer is one that tells how you feel.

Sentences About the Police

1. The police are pretty nice people.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

2. The police are brave.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

3. The police break the law.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

4. The police are a great help to people.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

5. The police enjoy kicking people around.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

go to page 3 ⟶

6. Without the police, there would be crime everywhere.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

7. The police are doing a good job.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

8. The police are dumb.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

9. The police are just as bad as the people they put in jail.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

10. The police are honest.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

11. The police are friendly.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

go to page 4 ⟶

12. The police are not needed in our country.
    ( ) Strongly Agree
    ( ) Agree
    ( ) Unsure
    ( ) Disagree
    ( ) Strongly Disagree

13. The police don't care what happens to you after they pick you up.
    ( ) Strongly Agree
    ( ) Agree
    ( ) Unsure
    ( ) Disagree
    ( ) Strongly Disagree

14. The police are always mad at kids.
    ( ) Strongly Agree
    ( ) Agree
    ( ) Unsure
    ( ) Disagree
    ( ) Strongly Disagree

15. The police care for young people.
    ( ) Strongly Agree
    ( ) Agree
    ( ) Unsure
    ( ) Disagree
    ( ) Strongly Disagree

16. Young people are treated fairly by the police.
    ( ) Strongly Agree
    ( ) Agree
    ( ) Unsure
    ( ) Disagree
    ( ) Strongly Disagree

17. The police stop and question people because it makes them feel powerful.
    ( ) Strongly Agree
    ( ) Agree
    ( ) Unsure
    ( ) Disagree
    ( ) Strongly Disagree

APPENDIX D

Survey Manual for the Finished Scale

Survey Manual

How Students Feel About the Police

(A survey of student attitudes toward the police)

Introduction

My name is Paul Tinsley, and I am a graduate student at Simon Fraser University, Faculty of Education. I am also a police officer with the Matsqui Police Service. Presently, I am engaged in educational research in this school district, and I am especially interested in the attitudes of students toward policing as it exists in a democracy.

I would very much appreciate your help in administering this survey. No names will be used, and the results will be scored in group form only. The results of this study are extremely important, as I am trying to assess the validity of the survey. The importance of your cooperation as the administrator of the survey cannot be overstated.

Should you have any questions, or would like additional information regarding the study or the results, please feel free to contact me (phone the Matsqui Police at 859-5225, or my home at 859-0466).

Thankyou.

Directions for Administration

As this is a survey of attitudes, the biggest problem is to obtain honest answers, as the students might have a tendency to respond to the statements in a manner they they feel is socially desirable.

Research has shown that this problem can be overcome through attention to a few details in administration, so I encourage you to follow these directions carefully.

1. You (the teacher administering the survey) are the most important factor. A teacher has a strong rapport with his/her students, and the students will most often respond as instructed.

2. Ensure that conditions are conducive to the administration, and that enough time is set aside (12 minutes for secondary students; 15 minutes for elementary students).

3. Assure the students that participation is voluntary; and for those who do participate, that the results are confidential. Anonymity is guaranteed.

4. Assure the students that this is not a test, that there are no right or wrong answers. What is requested is their honest answers.

5. Convince the students of the importance of the research, and the importance of their cooperation.

6. Read the directions found on page 1 of the survey aloud. Do the example with the class, and answer any clarifying questions.

7. Direct the students to complete the survey on their own, responding once to each statement.

Class Profile

(To be completed by the class teacher)

School:_____ Teacher:_____

Grade:_____ Racial:_____

Profile Nonparticipants:_____

_____

Intelligence:_____

_____

Academic Performance:_____

_____

_____

Behavioral/Attitude:_____

_____

Description of Area:_____

_____

Family Background:_____

_____

Anything Unusual:_____

_____

_____

_____

Thankyou!

APPENDIX E

Finished Scale

How Students Feel About the Police

* You are not asked to sign your name.

Age:_____ Grade:_____ Sex:_____

School:_____

Directions (To be read aloud to the class.)

This is not a test. There are no right or wrong answers. Just answer each sentence as honestly as you can.

Each sentence asks how you feel about the police. You are to tell how you feel by making an X by one of the answers found under each sentence.

Here is an example:

The police work hard.
( ) Strongly Agree
( ) Agree
( ) Unsure
( ) Disagree
( ) Strongly Disagree

Which answer tells best how you feel about the sentence? Put an X by only one answer.

Please read each sentence carefully. Remember, the best answer is one that tells how you feel.

Sentences About the Police

1. Young people are treated fairly by the police.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

2. The police are always mad at kids.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

3. The police are dumb.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

4. The police are friendly.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

5. The police care for young people.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

go to page 3 ⟶

6. The police are pretty nice people.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

7. The police are honest.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

8. The police are just as bad as the people they put in jail.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

9. The police <u>don't</u> care what happens to you after they pick you up.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

10. The police are doing a good job.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

11. The police enjoy kicking people around.
   ( ) Strongly Agree
   ( ) Agree
   ( ) Unsure
   ( ) Disagree
   ( ) Strongly Disagree

Bibliography

Adams, G. S. (1982). Attitude measurement. In H. E. Mitzel, J. H. Best, & W. Rabinowitz (Eds.), Encyclopedia of educational research (5th ed.) (pp. 180-189). New York: The Free Press.

Ajzen, I., & Fishbein, M. (1980). Understanding attitudes and predicting social behavior. New Jersey: Prentice-Hall.

Albrecht, S. L., & Green, M. (1977). Attitudes toward the police and the large attitude complex. Criminology, 15(1), 67-86.

Allington, R. L., Cramer, R. L., Cunningham, P. M., Perez, G. Y., Robinson, C. F., & Tierney, R. J. (1985). Focus: Reading for success (wheels and rockets) (teacher's ed.). Illinois: Scott, Foresman, & Company.

Allport, G. W. (1935). Attitudes. In C. Murchinson (Ed.), A handbook of social psychology (pp. 798-844). Massachusetts: Clark University Press.

Amoroso, D. M., & Ware, E. E. (1981). Adolescent's perception and evaluation of police. Canadian Journal of Behavioral Science, 13(4), 326-335.

Anastasi, A. (1976). Psychological testing (4th ed.). New York: Collier MacMillan.

Anderson, L. W. (1981). Assessing affective characteristics in the schools. Boston: Allyn & Bacon.

Baker, C. D. (1983). A 'second look' at interviews with adolescents. Journal of Youth and Adolescence, 12, 501-519.

Beckman, E. (1980). Law enforcement in a democratic society: An introduction. Chicago: Nelson-Hall.

Bewley, L. (January, 1986). Reformists? They belong in a cartoon. News & Views, 38(1), 5. (Published by the Metropolitan Toronto Police Association.)

Bible belt district or not? (1987, December 12). The [Abbotsford, Sumas, Matsqui] News.

Bird, C. (1940). Social psychology. New York: Appleton-Century.

Borg, W. R., & Gall, M. D. (1983). Educational research: An introduction (4th ed.). New York: Longman.

Bouma, D. H. (1969). Kids and cops. Michigan: William B. Eerdmans Publishing.

Bouma, D. H., & Williams, D. G. (1972, November). Police-school liaison: An evaluation of programs. Intellect, 119-122.

Brandstatter, A. F., & Radelet, J. A. (Eds.) (1968). Police and community relations: A sourcebook. California: The Glencoe Press.

Brenner, M., Brown, J., Canter, D. (Eds.) (1985). The research interview: Uses and approaches. Florida: Academic Press.

Chapman, A. W. (1956). Attitudes toward legal authorities by juveniles: A comparative study of delinquent and nondelinquent boys. Sociology and Social Research, 40(3), 170-175.

Chapman, A. W. (1960). Attitudes toward legal agencies of authority for juveniles: A comparative study of one hundred thirty-three delinquent and one hundred thirty-three non-delinquent boys in Dayton, Ohio (Doctoral dissertation, The Ohio State University, 1953). Dissertation Abstracts International, 1960, 20(7), 2943-2946.

Charles, M. T. (1980). The utilization of attitude surveys in the police decision-making process. Journal of Police Science and Administration, 8(3), 294-303.

Cronbach, L. J. (1946). Response sets and test validity. Educational and Psychological Measurement, 6, 475-494.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297-334.

Cronbach, L. J. (1970). Essentials of psychological testing (3rd ed.). New York: Harper & Row Publishers.

Cronbach, L. J. (1983). Course improvement through evaluation. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), Evaluation models, (pp. 101-115). Boston: Kluwer-Nijhoff.

Dale, E., & Chall, E. S. (1948). A formula for predicting readability. Educational Research Bulletin, 27(1), 11-20, 28.

Dale, E., & Chall, E. S. (1948). A formula for predicting readability: Instructions. Educational Research Bulletin, 27(2), 37-54.

Das, D. K. (1985). The image of American police in comparative literature. Police Studies, 8(2), 74-83.

Dawes, R. M. (1972). Fundamentals of attitude measurement. New York: John Wiley & Sons.

Dawes, R. M., & Smith, T. L. (1985). Attitude and opinion measurement. In G. Lindzey, & E. Aronson (Eds.), Handbook of social psychology (3rd ed.) (pp. 487-508). New York: Random House.

Derbyshire, R. L. (1968). Children's perceptions of the police: A comparative study of attitudes and attitude change. Journal of Criminal Law, Criminology, and Police Science, 59 (2), 183-190.

Doob, L. W. (1947). The behavior of attitudes. Psychological Review, 54, 135-156.

Edwards, A. L. (1957). Techniques in attitude scale construction. New York: Appleton-Century-Crofts.

Edwards, A. L, & Kenney, D. C. (1967). A comparison of Thurstone and Likert techniques of attitude scale construction. In M. Fishbein (Ed.), Readings in attitude theory and measurement (pp. 249-256). New York: John Wiley & Sons.

Eisner, E. W. (1983). Educational connoisseurship and criticism: Their form and functions in educational evaluation. In G. F. Madaus, M. Scriven, & D. L. Stufflebeam (Eds.), Evaluation models (pp. 335-347). Boston: Kluwer-Nijhoff.

Eisner, E. W. (1985). The educational imagination: On the design and evaluation of school programs (2nd ed.). New York: MacMillan Publishing.

Ellis, E. N. (1973). Evaluation of the involvement of a police officer in schools of the Killarney District (Report No. 73-14). Vancouver, B.C.: Board of School Trustees, Department of Planning and Evaluation.

Erickson, B. H. & Nosanchuk, T. A. (1977). Understanding data. Toronto: McGraw-Hill Ryerson.

Farley, J. (1981). Student interviews as an evaluation tool. Educational Leadership, 39, 184-86.

Finch, Jr., A. J., & Rogers, T. R. (1984). Self-report instruments. In T. H. Ollendick, & M. Hersen (Eds.), Child behavioral assessment (pp. 106-123). New York: Pergamon Press.

Fishbein, M. (Ed.) (1967). Readings in attitude theory and measurement. New York: John Wiley & Sons.

Frank, M. (1972). Modern English: A practical reference guide. New Jersey: Prentice-Hall.

Fry, E. (1968). A readability formula that saves time. Journal of Reading, 11(7), 513-515, 575-578.

Fry, E. (1977). Fry's readability graph: Clarification, validity, and extension to level 17. Journal of Reading, 21(3), 242-255.

Gall, G. L. (1983). The Canadian legal system (2nd ed.). Toronto: Carswell Legal Publications.

Gardner, P. L. (1975). Attitude measurement: A critique of some recent research. Educational Research, 17(2), 101-109.

Girod, G. R. (1973). Writing and assessing attitudinal objectives. Columbus: Charles E. Merrill Publishing.

Goran, S. P. (1970). Development of a scale of attitudes toward the police. Missouri: Missouri University, Columbia College of Education. (ERIC Document Reproduction Service No. ED 045 037)

Griffiths, C. T. (1982). Police school programs: The realities of the remedy. Canadian Journal of Criminology, 24(3), 329-340.

Gronlund, N. E. (1985). Measurement and evaluation in teaching (5th ed.). New York: MacMillan Publishing.

Gross, A. M. (1984). Behavioral interviewing. In T. H. Ollendick, & M. Hersen (Eds.), Child behavioral assessment (pp. 61-79). New York: Pergamon Press.

Guilford, J. P., & Frutcher, B. (1973). Fundamental statistics in psychology and education (5th ed.). New York: McGraw-Hill.

Guttman, L. (1944). A bases for scaling qualitative data. American Sociological Review, 9, 139-150.

Hess, R. D., & Torney, J. V. (1967). The development of political attitudes in children. Chicago: Aldine Publishing.

Horton, Jr., R. E. (1963). In H. H. Remmers (Ed.), Anti-democratic attitudes in American schools. Evanston: Northwestern University Press.

Jackson, D. N. (1982). Personality assessment. In H. E. Mitzel, J. H. Best, & W. Rabinowitz (Eds.), Encyclopedia of educational research (5th ed.) (pp. 1390-1398). New York: The Free Press.

Jayewardene, C. H. S. (1973). Police and the changing society: Problems, issues and solutions. University of Ottawa:

Department of Criminology.

Johnson, O. G. (1976). Tests and measurements in child development: Handbook II (1st ed., vol 2). San Francisco: Jossey-Bass Publishers.

Johnson, J., Hogan, R., & Zonderman, A. B. (1981). Moral judgment, personality, and attitudes toward authority. Journal of Personality and Social Psychology, 40(2), 370-373.

Kaplan, R. M., & Saccuzzo, D. P. (1982). Psychological testing: Principles, applications, and issues. California: Brooks/ Cole Publishing.

Kerlinger, R. N. (1964). Foundations of behavioral research: Educational and psychological inquiry. New York: Holt, Rinehart and Winston.

Klyman, F. I., & Kruckenberg, J. (1974). A methodology for assessing citizen perceptions of police. Journal of Criminal Justice, 2(3), 219-233.

Koenig, D. J. (1974). British Columbian's attitudes and experiences relevant to the police, law and crime. Victoria, B. C.: The Justice Development Commission Office of the Attorney General.

Koenig, D. J. (1975). Police perceptions of public respect and extra-legal use of force: a reconsideration of folk wisdom & pluralistic ignorance. The Canadian Journal of Sociology, 1 (3), 313-324.

Lapsley, D. K., Harwell, M. R., Olson, L. M., Flannery, D., & Quintanna, S. M. (1984). Moral judgment, personality, and attitude to authority in early and late adolescence. Journal of Youth and Adolescence, 13(6), 527-542.

Lemon, N. (1973). Attitudes and their measurement. London: B. T. Batsford.

Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 140, 3-55.

Loree, D. J. (1985). Police in a plural society. Canadian Police College Journal, 9(4), 391-412.

Lunney, R. F. (1986). Peace keepers or pencil pushers. Canadian Police Chief Newsletter, 5(9).

Marenin, O. (1983). Supporting the local police: The differential group basis of varieties of support. Police Studies, 6(3), 50-56.

Matsqui Economic Development Commission and the District of Matsqui Planning Department (1987, February). Matsqui Opportunity. (Available from The District of Matsqui, British Columbia)

Mattel, M. S., & Jacob, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. Educational and Psychological Measurement, 31, 657-674.

McGuire, W. J. (1985). Attitudes and attitude change. In G. Lindzey, & E. Aronson (Eds.), Handbook of social psychology (3rd ed.) (pp. 233-304). New York: Random House.

Meehan, K., Cathers, B., Badger, B., Blowers, T., Penner, W., Wray, R., Underhill, R., & Cassels, D. (1977 June). Report of the evaluation of the second year of the junior high police-school liaison program. Edmonton, Alta.: Edmonton Public Schools, Department of Research and Evaluation.

Moore, D. S. (1979). Statistics: Concepts and controversies (2nd ed.). New York: W. H. Freeman and Company.

Murphy, G., & Likert, R. (1967). Public opinion and the individual (reprint, 1938). New York: Russell & Russell.

Osgood, C. E. (1952). The nature and measurement of meaning. Psychological Bulletin, 49(3), 197-237.

Osgood, C. E., Suci, G., & Tannenbaum, P. (1957). The measurement of meaning. Urbana: University of Illinois Press.

Parker, W. C. (1984). Interviewing children: Problems and promise. Journal of Negro Education, 53, 18-28.

Pegler, M., & Wright, I. (n.d.). Attitudes toward the police: An experimental study. Surrey, B.C.: Royal Canadian Mounted Police, Surrey Detachment.

Peters, C. C., & Van Voorhis, W. R. (1940). Statistical procedures and their mathematical bases. New York: McGraw-Hill.

Pfiffner, J. M. (1969). The function of the police in a democratic society (series 7). University of Southern California: School of Public Administration Centre for Training and Development.

Phillips, J. L., & Coates, R. B. (1971). Two scales for measuring attitudes toward police. Wisconsin Sociologist, 8, 3-19.

Piliavin, I., & Briar, S. (1964/1965). Police encounters with

juveniles. The American Journal of Sociology, 70, 206-214.

Portune, R. L. (1965). An analysis of the attitudes of junior high school pupils toward police officers applied to the preparation of a work of juvenile fiction (Doctoral dissertation, University of Cincinnati, 1965). Dissertation Abstracts International, 1965, 5137.

Portune, R. L. (1968). The Cincinnati police-juvenile attitude project in police teacher curriculum development to improve police-juvenile relations, final report. (ERIC Document Reproduction Service No. ED 029 351)

Radelet, L. A. (1973). The police and the community. Beverly Hills: Glencoe Press.

Robinson, D. W. (1967). Police in the schools. Phi Delta Kappan, 48, 278-280.

Rokeach, M. (1973). The nature of human values. New York: Free Press.

Rundquist, E. A., & Sletto, R. F. (1975). Personality in the depression (reprint, 1936). Connecticut: Greenwood Press.

Rusinko, W. T., Johnson, K. W., & Hornung, C. A. (1978). The importance of police contact in the formulation of youths' attitudes toward police. Journal of Criminal Justice, 6(1), 53-68.

Salhaney, R. E. (1984). The police manual of arrest, seizure, and interrogation (2nd ed.). Ontario: Fransal Publishing.

Scott, W. A. (1968). Attitude measurement. In G. Lindzey, & E. Aronson (Eds.), Handbook of social psychology (2nd ed.) (pp. 204-273). Ontario: Addison-Wesley.

Seiler, L, & Hough, R. L. (1970). Empirical comparisons of the Thurstone and Likert Techniques. In G. F. Summers (Ed.), Attitude measurement (pp. 159-173). Chicago: Rand McNally.

Shane, P. G. (1980). Police and people: A comparison of five countries. St. Louis: C. V. Mosby.

Shaw, B. G., & Dobson, K. S. (1986). Cognitive assessment and depression. In T. V. Merluzzi, C. R. Glass, & M. Genest (Eds.), Cognitive assessment (1st paperback ed.) (pp. 361-387). New York: New York University Press.

Simmons, J. E. (1974). Psychiatric examination of children (2nd ed.). Philadelphia: Lea & Febiger.

Singer, M. S., Jonas, A., McCormick, I. A., & Walkey, F. H. (1985). Attitudes towards actors in the justice system:

Police and students in New Zealand. Police Studies, 8(1), 51-57.

Smith, J. K. (1983). Quantitative versus qualitative research: An attempt to clarify the issue. Educational Researcher, 12 (3), 6-13.

Southwick, R. A. (1974). Involvement of a police officer in schools in the Lord Byng District (Report No. 74-16). Vancouver, B.C.: Board of School Trustees, Department of Planning and Evaluation.

St. Pierre, P. (1981, Fall). The selling of the citizen-policeman. Symposium of Canadian Police Commissions. Quebec City.

Thornton, L. M. (1975). People and the police: An analysis of factors associated with police evaluation and support. The Canadian Journal of Sociology, 1(3), 325-342.

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

Thurstone, L. L., & Chave, E. J. (1929). The measurement of attitude. Chicago: University of Chicago Press.

Tittle, C. R., & Hill, R. J. (1967). Attitude measurement and prediction of behavior: An evaluation of conditions and measurement techniques. Sociometry, 30, 199-213.

Torney, J. V. (1971). Socialization of attitudes toward the legal system. Journal of Social Issues, 27(2), 137-154.

Triandis, H. C. (1967). Toward an analysis of the components of interpersonal attitudes. In C. W. Sherif, & M. Sherif (Eds.), Attitudes, ego involvement, and change (pp. 227-270). New York: John Wiley & Sons.

Triandis, H. C. (1971). Attitudes and attitude change. New York: John Wiley & Sons.

Tyler, R. (1973). Assessing educational achievement in the affective domain. Measurement in Education, 4(3), 1-8.

Weiner, I. B., & Elkind, D. (1972). Readings in child development. New York: John Wiley & Sons.

White, M. F., & Menke, B. A (1978). A critical analysis of surveys on public opinions toward police agencies. Journal of Police Science and Administration, 6, 204-218.

Wideen, M. (1987, July & August). Quantitative and qualitative approaches to evaluation. Lectures on curriculum evaluation, Simon Fraser University, Burnaby, B.C.

Wilt, G. M., & Bannon, J. D. (1976). Cynicism or realism: a critique of Niederhoffer's research into police attitudes. Journal of Police Science and Administration, 4(1), 38-49.