

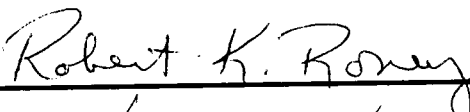
To the Graduate Council:

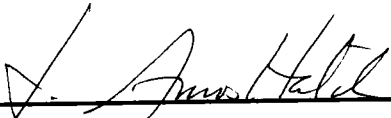
I am submitting herewith a dissertation written by Joseph B. Parker entitled "Tennessee Teachers and Testing: A Profile of Perceptions." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Education, with a major in Educational Administration and Supervision.

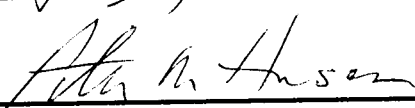


Dr. Gary Ubben, Major Professor

We have read this dissertation
and recommend its acceptance:







Accepted for the Council:



Associate Vice Chancellor
and Dean of The Graduate School

**TENNESSEE TEACHERS AND TESTING:
A PROFILE OF PERCEPTIONS**

A Dissertation

Presented for the

Doctor of Education

Degree

The University of Tennessee, Knoxville

Joseph Burr Parker

August 1994

Copyright © Joseph Burr Parker, 1994
All rights reserved

DEDICATION

This dissertation is dedicated to my wife

Patricia

the one person who stood by and offered
more than just moral support on more than one occasion.

Abstract

The purpose of this study was to determine the perception of the general level of knowledge that currently exists among teachers and administrators in the state of Tennessee about standardized testing, assessment, evaluation, utilization of standardized test results, and their attitudes toward standardized tests. Implicit in this purpose is the level of awareness of the science of tests and measurement and the level of understanding of testing in the education environment.

The subjects for the study were a randomly-drawn sample of teachers and administrators in the state of Tennessee. A survey instrument was devised through preliminary survey interviews and a pilot study and mailed to the respondents. The instrument was designed to address research questions specific to this study dealing with the use of student standardized test data by teachers and administrators in Tennessee, the current level of perceived understanding about basic testing and measurement terms and processes by professional educators in Tennessee, the training regarding tests and measurement teachers and administrators had received in their professional preparation, and the prevalent attitude towards using student standardized test data in teacher evaluations among practitioners in Tennessee.

The following elements emerged from the analysis of the data: the most frequent use of student standardized test data was for diagnosing student problems, limiting the effectiveness of the use of test data. The gap between educators' perceptions of testing terms and the use of those terms is large, indicating a possible lack of understanding. Information gained through various methods regarding tests and measurement was not utilized effectively by practitioners. Educators' attitudes toward standardized tests, especially in the case of the use of standardized test data as a component of teacher evaluation, were doubtful and contradictory, and, in many cases, negative.

TABLE OF CONTENTS

CHAPTER	PAGE
I. INTRODUCTION.....	1
Statement of the Problem.....	4
Purpose of the Study.....	5
Importance of the Study.....	6
Assumption.....	8
Limitation.....	8
Delimitation.....	8
Definitions.....	8
II. A REVIEW OF THE LITERATURE.....	11
Introduction.....	11
A Brief History of Testing.....	12
Uses and Purposes of Testing.....	14
Accountability.....	15
Curriculum Establishment.....	19
Decision-Making and Policy Formation.....	20
Criticisms of Current Educational Testing.....	21
Level of Knowledge of Current Practitioners.....	27
III. METHODOLOGY.....	33
The Population.....	33
Data Collection Procedures.....	33
Analysis of the Data.....	36

IV. PRESENTATION AND ANALYSIS OF THE DATA.....	40
Population Demographics.....	40
Data Analysis.....	54
Use of Standardized Test Score Data.....	54
Perceptions of Testing Terms and Term Usage.....	82
Previous Training in Psychometrics.....	103
Attitudes toward Standardized Tests, "Value-Added"	
Assessment, and Changes in Teaching Methods.....	111
V. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS.....	135
Summary.....	135
Conclusions and Discussion.....	143
Recommendations.....	147
BIBLIOGRAPHY.....	149
SECONDARY REFERENCES.....	154
APPENDICES.....	157
Appendix A - Survey Questionnaire.....	158
Appendix B - Superintendent's Letter.....	162
Appendix C - Initial Survey Cover Letter.....	163
Appendix D - Subsequent Survey Cover Letter.....	164
Appendix E - Responses to Open-Ended Question #26 or	
Other Comments.....	165
VITA.....	177

LIST OF TABLES

TABLE	PAGE
1 Age of Respondents.....	4 1
2 Years' Experience in Education.....	4 3
3 Grades Currently Taught.....	4 4
4 Highest Educational Level.....	4 6
5 Working on Advanced Degree by Age, Years' Experience, and Educational Level.....	4 7
6 A Comparison of Survey Respondents and Telephone Respondents by Age, Years' Experience, Grades Taught, and Educational Level.....	4 9
7 A Comparison of Fall Respondents and Spring Respondents by Test Score Use, Attitudes toward Standardized Tests, Attitudes toward Value-added Assessment, Term Perception, and Prior Training.....	5 2
8 Overall Use of Test Score Results.....	5 5
9 Use of Test Scores by Assignment.....	5 9
10 Use of Test Scores by Age.....	6 2
11 Use of Test Scores by Years' Experience.....	6 8
12 Discussion of Test Score Results with Others.....	7 4
13 Preferences for Bases of Educational Decisions.....	7 6
14 Bases for Educational Decisions by Years' Experience.....	7 8
15 Perceptions and Use of Testing Terms.....	8 4
16 Perceptions of Testing Terms by Age.....	8 7
17 Perceptions of Testing Terms by Years' Experience.....	9 3
18 Contexts of Term Use.....	10 2
19 Percentage of Use of Testing Terms in M-Team Meetings.....	10 4
20 Course Descriptions.....	10 5

LIST OF TABLES (continued)

TABLE		PAGE
2 1	Use of Outside Activities to Increase Knowledge of Testing.....	1 0 8
2 2	Frequency of Activities Related to Testing.....	1 1 0
2 3	Attitudes Concerning Standardized Tests.....	1 1 2
2 4	Attitudes toward Standardized Tests by Age.....	1 1 4
2 5	Attitudes toward Standardized Tests by Years' Experience.....	1 1 8
2 6	Attitudes Concerning Using "Value-Added" Test Scores as a Component of Teacher Evaluation.....	1 2 3
2 7	Attitudes toward "Value-Added" Assessment by Age.....	1 2 4
2 8	Attitudes toward "Value-Added" Assessment by Years' Experience.....	1 2 9
2 9	Degree of Change of Teaching Aspects.....	1 3 4

CHAPTER I

INTRODUCTION

In today's ever-changing educational setting, virtually all students in the nation's schools will have taken a standardized test of some sort by the time they reach the fourth grade, some as early as kindergarten. These students will be unwilling recipients of several high-stakes decisions made about their futures based on the results of these tests. Decisions on class placement, special service eligibility, grade retention, minimum-competency, graduation, or college entrance will be based in some part on a score on a standardized test.

As an example of the magnitude and pervasiveness of testing in this country, consider this statement made almost twenty years ago:

In 1974 the Commission on Tests of the College Entrance Examination Board stated in "The Importance of College Entrance Testing," that

... by about 1980 only one young American in ten will not graduate from high school; well over half--six, perhaps seven--of the other nine will go immediately to college; and virtually all of this college entry will be mediated in part by tests (Tyler & Wolf, 1974, p. 51).

Now, that number has grown such that today almost 90 percent of all students graduate from high school, and over 65 percent of them enter an institution of higher education of some sort. And, as predicted, almost all of that educational endeavor is based in part on the results of some form of a standardized test.

As yet another example of the magnitude of testing in this country, it has been estimated that 20 million standardized tests are given in grades K-12 each year (Lewis, 1986). *FairTest*--a consumer testing watchdog organization--estimates that the number is closer to 100 million. The total testing industry--including higher education and professional exams--is about a \$1 billion a year enterprise (Lewis, 1986).

Additionally, according to the *Christian Science Monitor* (August 23, 1985), more than 100 test publishers provided \$100 million worth of tests for K-12 education in one year alone.

To further illustrate the pervasiveness of testing, more than 40 states require minimum competency exams of students; 34 states require students to pass a test in order to graduate from high school; 32 states now require testing of prospective teachers; and three states underwent harsh criticism because of legislative mandates for testing of veteran teachers (Herman & Dorr-Bremme, 1984; Lewis, 1986). It is no wonder then, that testing has become a major industry and occupies much of educators' time, and, in some instances, a significant portion of their budgets as well.

Who administers these tests, and why are they given? In many areas, they are given by counselors, either state-employed or school personnel. In most cases in K-12 education, they are administered by a guidance counselor or classroom teacher. Many tests are given to satisfy state or district mandates to assess student knowledge and learning for program evaluation. Some critics feel that today's students are over-tested, that tests in today's society are misused, and that students do not and will not derive any benefits from increased testing (Darling-Hammond, 1991; Madaus, 1991; Shephard, 1991; Stake, 1991; McLaughlin, 1991). Others, including writers of reports such as *A Nation at Risk*, proclaim that test scores are decreasing nationwide and are the symptoms of decay. Some even wonder if test scores are reported correctly and have the same meaning for those who use them (Koretz, 1988; Stiggins, 1991).

Conversely, some defend the use of standardized tests on the grounds that they are effective ways of assessing the teaching, learning, or leadership that exists--or does not exist--in the nation's schools (Ashworth, 1990). Still others see test scores becoming part of a national movement toward educational reform (Lieberman, 1991; McLaughlin, 1991). A significant portion of this reform movement uses standardized test scores as a means for accountability: accountability for teachers, schools, and education in general.

Yet testing as a means of determining accountability in many states is still untried (Cornett & Gaines, 1991).

How is accountability measured? One method is to assess the performance of the worker against some standard and then reward or rehabilitate as appropriate. In education, the process becomes more complicated than simple comparison to some basic standard. To follow a factory model, if the output of the school is the student, then the student becomes the product by which the organization can be measured. For teachers, accountability means the responsibility of producing a product that will measure up in the marketplace. In other words, the teacher imparts sufficient information to the students so that they can demonstrate measurable gains in learning. Cook (1984) has suggested that the measurement of student achievements due to the teacher's performance is crucial to any teacher accountability system. Simply put, the teacher's accountability rests on the gains demonstrated by the student.

The use of student standardized achievement test data for teacher accountability has been proposed by the state of Tennessee as it embarks on a ambitious voyage with Governor McWherter's Basic Educational Program. The program provides, among other things, for a method of student testing for teacher accountability utilizing a new statistical analysis refined by two researchers at the University of Tennessee, Robert McLean and William Sanders. This analysis procedure proposes that a statistical methodology exists for evaluating teachers based on student standardized achievement data that considers such outside factors as environment and prior learning and then compensates for those factors to determine the information added to the student's knowledge by the district, school, or teacher. This provides its title, "value-added" assessment (McLean & Sanders, 1989, 1991).

The Basic Education Program (BEP) for the state of Tennessee declares, as part of its reform effort, that by April 1, 1993, school system effect on the educational progress of students in grades 3-8 is to be calculated; by July 1, 1994, school effect is

to be calculated; and by July 1, 1995, teacher effect is to be calculated, all based on value-added assessment results. In subsequent years, secondary school students and teachers will undergo the same process. The goal is to establish accountability for systems, schools and teachers that can be measured.

With so much at stake, it is not surprising that student testing and teacher accountability have become two of the most important issues in education today. What do teachers and administrators know about the testing that permeates the educational system in which they strive to produce the student who is ready to take his place in the world marketplace? What is the level of knowledge that exists in the educational arena by the participants, and what is its origin? Stiggins (1991) defines this level of knowledge as "assessment literacy," the informed consumerism of testing information and knowledge of results. As the topic of high-stakes student testing comes to be more closely associated with these professionals and their accountability, the issue of the level of awareness of student standardized testing information becomes one of concern.

Statement of the Problem

Teachers, principals, administrators, state boards of education, legislators, and others interested in education use the results of student standardized achievement tests to assist them in making the decisions required in today's increasingly complex society. The purposes for which testing data are used by those who guide educational policy and decision making are many and varied. It is imperative that those who make the important policy and curriculum decisions in the public schools of this country have the best information available. Student standardized test data are just one of the many kinds of the information available to them.

Standardized testing means little for those who teach the students and for those who must determine the policies that guide those teachers without some understanding of

the processes involved. Discussions with teachers and administrators and research on this topic reveal that many practitioners in Tennessee and elsewhere lack a comprehensive understanding of the science of tests and measurement as utilized by educational institutions and test manufacturers. If, as the Basic Education Program in Tennessee mandates, student test results are to become part of a teacher's evaluation, then it is imperative that teachers and administrators become as knowledgeable as they can about testing and the use of student achievement test results.

Purpose of the Study

The purpose of this study is to determine the perception of the general level of knowledge that currently exists among teachers and administrators in the state of Tennessee about standardized testing, assessment, evaluation, and utilization of standardized test results. Implicit in this purpose is the level of awareness of the science of tests and measurement, psychometrics. Also implicit is the level of understanding of the importance of testing in the education environment.

This study will attempt to answer the following questions:

1. How have student standardized test data been utilized by teachers and administrators in Tennessee?
2. What is the current level of perceived understanding about basic testing and measurement terms and processes by professional educators in Tennessee?
3. What training regarding tests and measurement have teachers and administrators received in their professional preparation?
4. What is the prevalent attitude towards using student standardized test data in teacher evaluations among practitioners in Tennessee?

Importance of the Study

At one time, proponents at the federal level called for national tests to assist in assessing the accountability of the nation's education establishment. While in office, both former President Bush and Education Secretary Lamar Alexander proposed as part of the America 2000 program that every student in the nation undergo voluntary testing. Currently, President Clinton and Education Secretary Riley have echoed President Bush's plan for reform by asking Congress to write into law the six goals that Bush and the nation's Governors agreed to in 1990. They do not, at this time, have any proposal for national testing of students or teachers other than imposing voluntary standards.

Some state governments have also entered the accountability arena, and some have mandated that tests be used to measure the effectiveness of their own institutions (Archbald & Newmann, 1988; Cohen, 1986; Ewell & Boyer, 1988). Business and industry have joined the accountability movement, citing the inability of schools to prepare their prospective work force for the rigors of the workplace (Haney & Madaus, 1986; Herman & Dorr-Bremme, 1984; Lewis, 1986; Lieberman, 1991; McLaughlin, 1991).

Educational testing is in a state of flux. Every state has some sort of testing policy in place. Many states, such as Oklahoma, Kentucky, and South Carolina, have tied testing of students to reform measures (Cornett & Gaines, 1991).

Operating through the State Department of Education (SDE), the state of Tennessee has stipulated in its BEP that every local education agency (LEA) devise performance goals for itself. If those goals are not met, and if their academic performance gains do not measure up, the state may place those LEAs on probation. The LEAs will be evaluated in such areas as fiscal accountability, value-added assessment, test score comparisons with state and national norms, rate of progress within each grade and academic course, school attendance and dropout rates. Minimum criteria have been

established for evaluating certificated employees to include classroom observation, review of prior evaluations, personal conferences, and results of value-added assessments (*Education Improvement Act of 1992, 1992*).

With increasing amounts of money being requested by educational institutions and more and more resources being allocated to the testing of students, a need exists for additional and better information that will enable teachers, administrators, legislators, and current educational policy makers to make the best decisions possible in the areas of testing and assessment. With the passage of the Basic Education Program (BEP) of 1992 in the state of Tennessee mandating value-added assessment for teachers, a need exists to determine the level of understanding of testing procedures and uses by those who will be evaluated by the results of those same tests.

Before a teacher is evaluated using value-added assessment data, it is imperative that the teacher understand what will be included in that assessment. This may call for adding courses in tests and measurement at the undergraduate level in teacher preparation programs and requiring such courses at the graduate level in professional programs. In addition, in-service opportunities should be made available to those practitioners in the field (Johns & Davis, 1991; Stiggins, 1991; VanLeirsburg & Johns, 1991).

The use of standardized achievement testing data as a means of evaluating teachers may be a step in the right direction. Or, it may, as some critics point out, take away the time needed for schooling by focusing instruction on the test itself. The effects on students, teachers, administrators, legislators, and those involved with educational decision making have yet to be realized by such a move toward increased testing. Methods other than strict high-stakes testing might be used as a component of teacher evaluation, but that does not negate the need for teachers and administrators to know and be aware of the complete process--to be, as Stiggins (1991) suggests "assessment literates" rather than illiterates (p. 535).

Assumption

It is assumed, for the purposes of the study, that professional educators in the state of Tennessee will have undergone a professional training program in an accredited teacher preparation institution.

Limitation

The study is based on the perceptions of the survey population, not the actual knowledge level. Due to the nature of the method of the study, actual knowledge level could not be ascertained without the use of a different form of evaluation. Consequently, only an estimation of test literacy is presented.

Delimitation

This study was delimited to a random sample of teachers and administrators in the public schools of the state of Tennessee.

Definitions

For the purposes of this study, the following definitions were used:

adjusted score : This is a score that has been altered through some statistical analysis, such as the analysis of covariance, to equate for initial differences on some variable. Such treatment is undertaken to assure the score is a result of treatment.

assessment : While synonymous with evaluation, this is also the method used to refer to the diagnosis of an individual's problems.

criterion-referenced test : Such tests give the estimate of the proportion of questions in a domain that a student will answer correctly, i.e. one that measures whether a learner has mastered a particular type of intended knowledge.

evaluation : This is the systematic process of collecting and analyzing data upon which decisions may be made.

grade equivalents (GE) : These are scores on a test that are equal to the median score for the norming group. If a student obtains a score on a test that is equal to the median score for all the beginning (in September) sixth-graders in the norm group, then the student's GE is 6.0.

mean : The arithmetic average of a set of scores is called the mean.

measurement : This is the use of observations, rating scales, or any other device that allows us to obtain information in a quantitative form; it can refer to both the score obtained and the process used to obtain that score.

median : This is the point in an ordered series below which 50 percent of the scores lie.

mode : This is the most frequently occurring value in a set of scores.

normal distribution : Called the familiar "bell-shaped" curve with the mean at 50 percent, the scores are arranged symmetrically around the mean.

norm-referenced test : A test whose scores have been referenced to a norming group, those scores of subjects in a reference (or normal) group, is referred to as a norm-referenced test.

percentile : This is a point on an ordered score distribution below which a certain percentage of the scores fall.

percentile rank : This is a person's, or score's, relative position on a normal distribution, or the percentage of students' scores that fall below that individual's obtained score.

raw scores : These are scores in their original form before they have undergone any form of transformation.

scaled score : These are scores that have undergone some mathematical transformation to force them into a particular interval.

standard score : They are transformed scores for which the resulting set of values has a distribution shape identical to the original raw-score distribution with a mean equaling 0 and a standard deviation equaling 1.

standardized tests : Such tests are commercially prepared by experts in measurement and subject matter, administered under uniform procedures, and have been previously administered to a norming group to obtain performance norms for comparison purposes.

stanine : Also called standard nine scores, these are derived scores with a mean of 5 and a standard deviation of 2; they are interval scores represented from 1 to 9 that divide the baseline of a normal distribution into nine equal divisions.

statistical process control : The use of measurement and data to adjust and continuously improve the process of controlling the statistics that are used when transforming raw scores into scaled scores is a form of total quality management.

test : This is the presentation of a standard set of questions to be answered; a method of evaluation of individuals by asking questions.

CHAPTER II

A REVIEW OF THE LITERATURE

Introduction

A review of the literature of testing and educational assessment in the United States reveals one salient fact: testing and measurement is one aspect of the educational milieu that has found an important resting place and will not be dislodged. To indicate the growth of testing in the educational arena, consider the following chronology. A survey conducted in 1967 by the Educational Testing Service (ETS) showed that there were 74 state testing programs in 42 states, with 18 states offering 2 or more programs (Dyer & Rosenthal, 1974). A 1976 report by the National School Boards Association found that approximately three quarters of school districts in the nation conducted some sort of testing program for its students (National School Boards Association [NSBA], 1977). Three years later, the National Education Association (NEA) found in its study 43 states that had a statewide testing program, an increase of four states in only three years (Norman, 1980). In 1984, the Center for the Study of Evaluation (CSE) at the University of California at Los Angeles observed that more than 40 of the 50 states had some sort of minimum competency testing program for its students (Herman & Dorr-Bremme, 1984). In a later study from ETS, all 50 states were found to have some sort of testing activity (Goertz, 1988). Of the 40 states that operate some sort of high school competency testing program, 19 use their tests as a requirement for graduation from secondary school (Jaeger, 1991). Not only has the use of testing in this country grown until it now encompasses every state in the union and the United States protectorates, but successful performance on the tests has become required for the completion of many programs.

Along with this increase in testing comes an accompanying increase in criticism from those who state that testing is being overused, as well as misused, to the students' detriment. Some strongly aver that research evidence from nearly 20 years of testing shows that testing policies have not had the positive effects that were intended when they were implemented; rather, they have had the opposite effect, unwanted and unintended negative effects (Lieberman, 1991). Others look at the commercial influence exerted by test publishers and non-school agencies involved in the testing industry and decry the cheapening of students' educational opportunities (Darling-Hammond, 1991). In this same vein, George Madaus, Director of the Center for the Study of Testing, Evaluation, and Educational Policy, offered this observation about the nation's move toward national tests:

Before lurching down the road to national testing, we must be clear about why we will test, what we will test, and whom we will test, as well as about the form of the test, the use of the results, the fiscal and educational costs of testing, and the nature of the infrastructure that must be created to develop and administer the tests and to control the testing program (1991, p. 227).

Clearly, questions and concerns about testing in this country are in the forefront of many minds. For over 20 years, almost every state in the union has imposed some form of uniform, mandatory testing on its school systems, its schools, and its students (Jaeger, 1991). Testing, whether for promotion in grades, placement in programs, measurement of achievement, or assignment of accountability is here to stay. Whether that testing will produce the desired results remains to be seen.

A Brief History of Testing

Testing, the method of evaluating individuals by questioning, has existed for centuries. The use of testing and measurement for some evaluative purpose has been documented as far back as 4000 B.C. In China written examinations were required in

the civil service system to assess just how much certain Chinese civil servants had learned (Linden & Linden, 1968; Norman, 1980). Other countries also had methods for ascertaining the learning of its students. The ancient Greeks, noted for the importance they attached to academics, engaged in the Socratic method for extending learning. During the Middle Ages, the University of Paris introduced the use of the oral examination for its master's degree candidates. This process then spread over Europe and became a common occurrence in almost all European universities (Linden & Linden, 1968).

From Alfred Binet's multiple-student assessments for intelligence quotients to the measurement of a single individual by Sir Francis Galton, testing has undergone a rapid and ever-changing process as it moved across the Atlantic to the United States. Perhaps most remembered for its uniqueness, the psychological testing in World War I developed by the U. S. Army was used to select those who could quickly learn to be officers and technical personnel. These tests were so effective that they impressed not only the psychologists of that time, but also leaders in education, business, and civic affairs. Schools, colleges, and large industrial organizations were seen as the proper civilian settings for the initiation and development of testing for the selection and sorting of school students (Tyler & Wolf, 1974).

From that beginning, the principles and practices of educational testing first begun in 1918 have continued until today with very little change. Testing was adopted and developed as a tool for performing the functions of grading, classifying, and other sorting functions; schools at that time were not particularly interested in helping students learn (Tyler & Wolf, 1974). Early aptitude tests were used for initial sorting and indicated for both teachers and parents the likely success of children in the educational system. Later this term, aptitude, was applied to examinations that were devised to aid in preliminary classification and guidance. Achievement tests were

developed to indicate a student's progress in school as compared to other students (Tyler & Wolf, 1974). The states themselves began testing of students along the same lines that the Army had used but for different purposes. Rather than selecting officers and technical personnel, the schools were primarily interested in achievement.

According to Henry S. Dyer and E. Rosenthal (1974), there were three major events in the middle 1960s that occurred nationally that affected the ways of thinking about educational assessment by the states and their school districts. The first, in 1964, was the formation of the Exploratory Committee on the Assessment of Progress in Education. This committee eventually became the National Assessment of Educational Progress (NAEP) in 1969, which is the national assessment program utilized currently. The second, in 1965, was the enactment of the Elementary and Secondary Education Act which required that school systems assess by objective means the effects on student achievement produced by federally funded programs for the educationally deprived. The third, in 1966, was the publication of the Coleman report on *Equality of Educational Opportunity*, which attempted to assess in terms of measured pupil achievement, the quality of service the schools were supplying to various segments of the population. These three events shaped the form and method of testing that exists to the present day. Without them, testing would not have evolved in this country to its present state.

Uses and Purposes of Testing

The purposes of testing are many and varied. As previously stated, the main use of testing seems not to have been only for the student's benefit. For example, in Herman and Dorr-Bremme's 1984 study, eight common uses for tests were found. Those with the greatest importance were: (1) curriculum planning, (2) communicating to parents, and (3) reporting to the district. Those next in importance were: (4)

informing the public, (5) student promotion, and (6) assigning student to classes. Those least in importance were: (7) teacher evaluation and (8) allocating funds (p. 43). Here, student concerns appeared in the middle of the spectrum, not at the top. Herman and Dorr-Bremme found that test scores made their greatest contribution in curriculum planning; the least contributions were made in teacher evaluation and fund allocation.

When scores are used for the student, their uses include tracking, retention in grade, graduation, and rewards and sanctions (Darling-Hammond, 1991). Linden and Linden saw a significant additional function of measurement as diagnosis: "to determine the nature and causes of unsatisfactory adjustment to the school situation for the purpose of correcting and preventing maladjustment" (1968, p. 80). Madaus (1991) agreed with Linden and Linden but added another concern:

Historically, testing in the U.S. has been used to diagnose and classify students and to assign them to educational treatments. For years test scores have been used to place students in different curricular and instructional tracks, to retain them in grade, and to assign them to remediation, readiness, and special education programs. The potential benefits of using tests to place students in these treatments have been taken for granted. The associated harms went unnoticed for a long time; they simply are not as direct, visible, physical, or dramatic as examples from medical technology. Nonetheless, they can be devastating for many students (p. 227).

This is a major concern of those who think that testing should be more firmly regulated.

Accountability

Mandated, standardized national testing is now being discussed as a means of improving education by holding schools accountable. . . . The tests are supposed to measure the most important outcomes of schooling--those for which the education system should be held accountable--while providing direction and motivation for teachers and students. They will become the standard by which the public can measure success or failure. (Lieberman, 1991, p. 219)

Accountability in education appears to fall under two broad umbrellas: first, accountability for education itself, and then for its practitioners. As Lieberman suggested, the measure of accountability is thought to rest upon test scores. This is a heavy burden for test scores to bear.

McLaughlin (1991) suggested that there are five important themes that emerge from a discussion of testing and accountability. He argued that it becomes important what is measured, that testing for testing's sake does not a measurement make. If we intend to tie accountability to testing, then we must ensure that we test what we need to test. Second, he suggested that standardization should not be confused with standards. As we rush toward the use of the norm-referenced test, we tend to be overcome by the "Lake Wobegon" effect, where we strive to put all our students above average. That is a fallacy, but one that is prevalent in present educational circles (Koretz, 1988).

Third, he suggested that tests constitute a limited lever for reform, that using test scores as a means of moving educational reform ideas forward is less a mover than it appears. Test scores in and of themselves only mean something when meaningfully compared to something else. Fourth, test-based accountability plans often misplace trust and protection in that one's belief in test scores are only as good as the test and the protection that the test data imply. Undue reliance on too much information can make doubting Thomases out of even the most hardened believer. Finally, he argued that the process of setting standards is as important as the standards themselves. To make a standard meaningful, the process by which that standard is set must be as stringent as the mark. To do less cheapens not only the goal, but the process used to attain the goal.

Teachers worry about such accountability systems primarily because they feel that there are too many outside influences that affect students' work and to hold them accountable for such factors is really unfair. Also, teachers feel that such accountability systems, those relying on test results, will drive out the very kind of instruction that is

most effective with the low-achieving or nontraditional students. Missing will be the type of instruction that concentrates on concepts, discussion, and active student participation--those things that tests cannot measure effectively. Tests tend to concentrate more on facts, memorization, and the passivity of the student (McLaughlin, 1991).

Wiggins (1989) suggested that standardized testing came about and grew because the school transcript became untrustworthy. An "A" in one course meant only that the adult in charge of the course felt that that student's work deserved an "A." But what was the standard? Compared to whom? To what? The high school diploma, tied to no other standard than credits earned, provided no really useful information about what students studied or what they actually did with whatever it was they studied. Wiggins (1989) further suggested that if schools do not rethink their requirements, their criteria for awarding of diplomas, then "standardized test makers will continue to succeed in hawking simplistic, norm-referenced tests to districts and states resigned to using them for lack of a better accountability scheme" (p. 42).

Ron Brandt (1989), in his conversation with George Madaus, elicited this observation from Madaus:

When important things like graduation, promotion from grade to grade, teacher evaluation, school district certification, and financial support--major aspects of accountability affecting teachers or administrators--depend directly on test scores, it corrupts the process you're trying to improve. . . . When the stakes are high, people are going to find ways to have test scores go up. That's true of any social indicator (p. 26).

What then is suggested about test results as accountability measures? An incompatibility exists between measurement to serve accountability and policy goals and measurement for instructional purposes. Shepard (1989) reiterated N. S. Cole's viewpoint in her argument of why we need better assessments. Cole suggested that the large-scale assessments must be formal, objective, time-efficient, cost-efficient,

widely applicable, and centrally processed. Most important, the results must be in a form useful to policy makers which usually means that they are reduced to an easy-to-read, easy-to-explain single score. The assessments designed to assist the instructional programs are informal, teacher-mandated, adapted to local context, locally scored, sensitive to short-term change in students' knowledge, and meaningful to students. They provide immediate, detailed and complex feedback, and they incorporate tasks that have instructional value in themselves. The information provided is in a form much more usable by those who can understand it and apply it to the task.

As far as using test scores for teacher evaluation and accountability, Herman and Dorr-Bremme (1984) found that "only 12% of the elementary-school principals and 11% of those in secondary schools say that they regularly or frequently set test-score goals for their teachers to meet or consider test results in teacher evaluation. . . . principals simply do not deem it appropriate to assess teachers' competence on the basis of their students' test performance. Most rely on their own observations of teachers' work in the classroom for this purpose" (p. 67).

Evaluating teachers by test results seems not to be the usual case. South Carolina, in its now-defunct District and Campus Incentive Programs, used student achievement test scores to determine monetary awards for districts and schools that displayed better-than-expected gains in student achievement. Oklahoma gave permission for two districts to begin merit bonus programs for teachers where one of the criteria for awarding of the bonus was the improvement of student achievement test scores. Tennessee, in its 1992 education reform act, proposed to do the same with its "value-added" assessment. This concept, based on the work of William Sanders and Robert McLean (1984; 1991) of The University of Tennessee, asserted that teacher evaluation using student achievement data is not only possible but very feasible.

Curriculum Establishment

Another widely-discussed aspect of the use of test scores is the concept of the test-driven curriculum. Some educators fear that if a national or state-mandated test were to be established for accountability purposes, it would drive the teachers to "teach to the test," thus undermining the very reason for testing. Darling-Hammond (1991) found that during the ten years when many test-oriented accountability measures were instituted in American schools, the use of teaching methods directed toward the higher-order thinking skills were neglected.

Between 1972 and 1980, public schools showed a decline in the use of such methods as student-centered discussions, the writing of essays or themes, and research projects or laboratory work. This should not be surprising. As evidence from many studies indicates, when high stakes are attached to scores, tests exert a strong influence on "what is taught, how it taught, what pupils study, how they study, and what they learn" (p. 222).

Others agree. George Madaus suggested that testing in this manner focuses the teachers' attention on the objectives to begin with, but as consequences increase, such as making an impact on their evaluation or promotion, the testing will become the curriculum (Brandt, 1989). Madaus also suggested that while it is the easiest thing in the world to follow a test-driven curriculum, in the long run it tends to "deprofessionalize" the practitioners by taking the important decisions out of their hands.

Shepard (1989) stated that there is much evidence that testing shapes instruction. Studies show that when the stakes of the testing are high, teachers take time away from areas such as science and social studies to devote more time to tested math skills. Tested areas are taught at the expense of untested areas, minimum proficiencies dictated by competency tests overwhelm class time, and within the bounds of test-driven content, there is a "dumbing down" of instruction as teachers teach the content of tests rather than the underlying concepts. She suggested that " . . . teaching to the test

cheapens instruction and undermines the authenticity of scores as measures of what children really know" (p. 5).

What then would a national test do to the local curriculum? That is a question that has not yet been answered. Maeroff (1992) suggested that a national system of assessment be developed as a companion to a national curriculum, as is being done in Great Britain, as a means of attempting to force the achievement of all students upward. One response to this push for a national test is the examination of alternative means of assessments now being scrutinized by some schools and states in the United States. South Carolina, in its Education Improvement Act (EIA) of 1984 has amended that act to include what is now called "Target 2000," no doubt in keeping with former President Bush's call for "America 2000." South Carolina's attempt at alternative assessment called for portfolios, performances, exhibits, essays, and other forms of assessment (South Carolina State Board of Education, 1990).

Decision Making and Policy Formation

Among the main purposes of test data are those of decision making and policy formation. Mehrens and Lehmann (1987) saw the functions of standardized tests as aids in decision making. Herman and Dorr-Bremme (1984) agreed that curricular programs and instructional strategies might be changed by virtue of a program evaluation that includes the use of test scores as measures of program effectiveness, i.e. test scores might influence student placement decisions. Tyler and Wolf (1974) saw test results affecting decisions made by administrators in the allocation of resources. They suggested that those with the best scores get the best resources, a not unheard of situation. Along these same lines, Cohen (1986) saw tests and the resulting data being required both to assess the effectiveness of the educational system in meeting certain goals, and to describe the nature of current educational practices, processes, and

programs which affect the attainment of those goals. In doing so, test results play an extremely important role in which decisions and what kind of decisions are made.

This line of reasoning leads to test score use for aiding policy development. As more and better data became available, legislators and government leaders utilized that data to help them formulate policies to help improve or reform education. The passage of the "21st Century Education Reform Act" in Tennessee was an example. Using a new test results evaluation methodology, Tennessee hopes to put in place a teacher evaluation system unlike others (*Education Improvement Act of 1992*, 1992). Cohen (1986) suggested that the policies that result from an action such as this are intended to mobilize, and maintain political support for reform. Former President Bush and Secretary Alexander vigorously promoted an educational reform movement. To sustain such a movement, states will have a need to collect data which proves that their recent investments in education are indeed paying off.

Finally, one last purpose of testing is that it can improve educational quality because it can focus attention on important goals, outcomes, and processes in education. According to Cohen (1986), it can set performance standards and sanctions to have a powerful local effect on the attainment of those goals. It can provide data about school incentive programs, as has been seen in South Carolina and Florida, and, finally, it can assist in assisting states in their regulatory functions such as determining accreditation standards as in Mississippi, Indiana, Vermont, and West Virginia.

Criticisms of Current Educational Testing

In any issue as broad or as important as this, criticisms abound. In fact, more critics of testing and the use and misuse of test scores exist than do those that praise them. Most of the criticisms of utilizing test scores have admonished against using test scores for the wrong reasons. Others have cautioned against maintaining a wholesale

reliance on test scores as the be-all or end-all. Still others have stated that testing is not the universal panacea that some believe it is. A review of the criticisms of utilization of testing revealed six major areas of criticism.

First, when political pressure and media attention attach high stakes to test results, scores can become inflated, thus giving a false impression of student achievement. One major critic of the use of tests and the reliance people place on them is Gerald W. Bracey. Writing in several journals, he has criticized the use of tests as indicators of educational progress--or lack of it. Bracey (1992) stated that ". . . there is no meaningful difference between ability and achievement as measured by paper-and-pencil tests. All tests measure only what students know and can do now; they do not measure 'potential.' Nevertheless, many people act as if they do" (p. 414). In attempting to measure what sort of "potential" a student might have, tests do no more than measure what the student knows at that particular time, on that particular day, in that particular room, in that particular situation. Assigning more information to a test score tends to devalue the student, not the test. More emphasis is placed on the test and its resultant score than on the person.

It has been suggested that the recent rise in test scores, which has been widely touted, has been due rather to some questionable activities on the part of teachers and administrators than to an actual increase in performance by students. The pressure that exists in some schools to increase scores has caused these schools to make their curriculum match the test, to emphasize test-taking strategies over the curriculum, and to even engage in some form of cheating (Bracey, 1991). The recent publication of what has been called the "Lake Wobegon" effect has added to this (Koretz, 1988). In Lake Wobegon, a fictional area created by humorist and social critic Garrison Keilor, all the children are "above average." Consequently, a false impression of the status of educational attainment is fostered.

A second major criticism dealt with the tendency of high-stakes test to narrow the curriculum. Ralph Tyler (1974) suggested that using tests that are designed to measure individuals in terms of deviations from the population mean results in just a abstract number with no concrete reference; it really has no bearing on reporting what tasks or skills the student can or cannot do. He felt that the use of ability grouping based on test scores can easily lead to a rigid "teaching system." This is much the same as was seen earlier where "teaching to the test" was felt to be the norm when the test begins to drive the curriculum.

Herman and Dorr-Bremme (1984) felt the same. They saw teachers taking time away from the regularly used curriculum to take time to "teach to the test." In informing when the tests will be given, explaining what they will cover, mentioning the testing format, or instructing in general test taking strategies, teachers suspend the on-going teaching of concepts and instead use class time to review the skills and content they know the test will cover. In doing so, they tend to negate the purpose of the test itself which should be the probing and identifying of skills and concepts the student knows from his exposure to that material in the classroom. As a result, the scores obtained from these types of classroom tests are likely to be skewed or false.

Third, high-stakes testing misdirects instruction toward the basic skills, neglecting the higher-order thinking skills so highly sought by educators. In this vein, Darling-Hammond (1991) suggested that because of the way that tests are written, they tend to ignore the higher-order thinking skills associated with true academic endeavors. By sitting at a desk filling in little bubbles, circles, or bars, the test puts the student in a passive role, not the active one teachers expect to see their students engaging in as they question, discuss, analyze, synthesize, and learn. Research on human learning processes and performance has suggested that many tests fail to measure the student's higher-order thinking skills or to involve them in real-world problem solving. Indeed, the

major emphasis of testing in this country today has seemed to be moving toward the testing of basic skills to the detriment of independent thinking. Many professional organizations and interested parties such as officials of the National Assessment of Educational Progress (NAEP), the National Research Council (NRC), the National Council of Teachers of English (NCTE), and the National Council of Teachers of Mathematics (NCTM) have said that the current decline in assessments of higher-order thinking skills is attributable to an emphasis on teaching basic skills (Darling-Hammond, 1991; Shepard, 1989).

Fourth, the kind of drill-and-practice instruction that tests reinforce is based on outmoded learning and testing theory. As testing has become more and more popular, so has the use of the norm-referenced test. Airasian and Madaus (1974) suggested that educational measurement did not develop along the lines of the physical sciences, but adopted instead a psychological model based on the concept of individual differences. "The psychological lineage, with its emphasis on individual differences, normal distributions, predictions, and the like is a primary reason norm-referenced (instead of criterion-referenced) measurement continues to dominate educational testing to this day" (p. 75). The concept of norm-referenced rather than criterion-referenced has been the one accepted in this country. We do not tend to measure a student's ability against some stated criteria, but against his peers. This has been what has caused a decline in test scores nationally (Bracey, 1991; Wiggins, 1991).

Because of the time required and the sheer number of students that must be tested, the multiple-choice format is used almost exclusively. This format also tends to de-emphasize the higher order thinking skills by negating any original thinking on the part of the student. Some have argued that with this format, students are rarely asked to develop an original answer or display an ability to organize relevant material into a coherent argument (Shepard, 1989). This too, has lent credence to the idea that tests

are based on learning theory that has become stultified and stagnate. Bracey (1991) suggested that such a strong use of the multiple-choice questions and responses so favored by test makers has so dominated the field of testing that many believe they have shaped the form of instruction. By following this format, testers have relied too much on the "passive" performance of the student thus not revealing a true indicator of the student's performance or knowledge level.

Shepard (1989) agreed, suggesting that other format problems have arisen in standardized tests that bear this out. Reading passages used by test writers to assess a student's comprehension or ability to understand the main idea are shorter than the student is familiar with. The shorter passages allow more questions to fit into the time constraint, but it also tends to produce a smaller number of item types. This becomes relevant when a student may appear to know a concept or skill when it is presented and measured in one format but appears not to be able to distinguish that same concept or skill when it is presented in a different format.

Fifth, because of pressure to achieve high test scores, more hard-to-teach children are rejected by the system. Lieberman (1991) suggested that testing limits students' opportunities to learn because it is used merely as a means to track and retain students. Practices such as these have been shown to lower achievement rather than raise it. These practices also tend to lower self-esteem and raise dropout rates. In doing so, testing tends to widen the gap that exists between white students and minority students and also between students from high-income families and students from low-income families. Madaus (1991) concurred, suggesting that there is evidence that competency tests for graduation and test-based retention policies have added to the increasing dropout rate rather than subtracting from it. This evidence has suggested negative consequences associated with testing rather than positive ones.

In addition to these criticisms, Shepard (1989) has proposed that today's

standardized tests are not the reasonable indicators of student learning they once were considered to be. The political climate has shaped the testing of students into an entity that has no relation to the outcomes derived from it. In "negotiating" the content of the tests, test writers, state curriculum experts, and teachers have agreed on what is to be included. When textbooks become part of the process, the content is weakened still further, leading to the claims that the tests do not measure the curriculum. Also, basic skills are emphasized more than higher order skills.

Linden and Linden (1968) offered this extensive comment on the criticisms of testing:

Critics who charge that testing will lead to less, rather than to more, humane treatment of individuals are concerned without cause. Each person is truly a unique and complex individual. To study him, therefore requires complicated ideas, complicated instruments, and complicated procedures. However, tests, like all other tools, which man has invented, can be and are being misused. There are those who forget, or fail to understand, the probabilistic nature of inferences derived from test scores and view such scores to absolute indices of human attributes. The problem is one of education, i.e., restricting the use of tests to those persons who are willing to take the trouble to understand them and to use them in an informed fashion. Our knowledge of testing today is vast. However, the basic concepts, even though complex, are neither too many nor too difficult to learn for persons properly motivated. Hope lies in research rather than invective, in careful evaluation instead of emotional out burst, and in a realization that while we have come a long way, there is an even greater distance to go (p. 91).

Maeroff (1992) in his examination of the English educational systems suggested that absence of the type of commercially produced tests that are so widely used in the U. S. indicates a philosophy of education that seems to be more humanistic than the one that exists here in this country.

With all the criticism that exists, testing is still the method of choice for every state department of education in this country for assessing student's skills. Even in the face of mounting evidence, state departments and the national Education Department have still called for increased testing of students. The most recent call for accountability has

just added even more fuel to this fire.

Level of Knowledge of Current Practitioners

What is the state of understanding of those who are current practitioners of the education process? Experience has shown that many teachers and administrators do not fully understand the testing process or what the scores really mean. Knowledge of the topic is generally vague and unclear. A teacher should be able to explain to parents what their child's test scores mean without ambiguity. Yet some cannot do that. Most parents have no understanding of psychometrics. It should be the job of the professional to explain what the scores indicate to those who need the information the most: the parents. The most recent *Gallup/Phi Delta Kappa* poll of the public's attitudes toward the public schools indicated that 79% of the respondents felt that standardized national tests should be used "to identify areas in which teachers need to improve their teaching skills" (Elam, Rose, & Gallup, 1992, p. 47). Educators should know how to use test data for improvement as well as for parent explanation. A review of the literature indicates that this does not seem to be the case.

While it seems appropriate that colleges of education should include courses in tests and measurement in their undergraduate and graduate teacher training programs, it appears that many do not. Roeder (1973) hypothesized that when elementary teachers graduated from college, they were more qualified to "conduct impromptu art and music lessons than they are to evaluate pupil performance" (p. 141). His statement was based on years of observing elementary teachers use standardized test data and other informal methods to evaluate their students. To test his hypothesis, he surveyed 940 accredited institutions of teacher preparation, asking how many credit hours were required in various areas before the teacher could become certified. Among the areas queried were evaluation, geography, art methods, music methods, physical education,

and "religion." He found that almost 70% of the institutions required less than two semester hours in the area of evaluation; 57.6% required none at all.

Olejnik (1979) attempted to address this issue by questioning a group of what he termed "non-measurement" experts, those being defined as "those directly involved with educational tests and measurement who do not have a degree related to this field of specialization" (p. 3). For his study, those who fit the definition included school administrators, counselors, and teachers in a single school district where an extensive achievement testing program was already in place. While he found that teachers and administrators across all grades seemed to view standardized achievement tests in a positive manner, he also found that test scores were used most only in the elementary grades. That may have had to do with test score accessibility, but it indicated strongly that the majority of usage by teachers and administrators came in the lower grades rather than the higher ones. A large percentage indicated that they were unfamiliar with test score interpretation, and almost three quarters of his respondents indicated they would welcome additional training in that area.

Ruddell (1985) examined the knowledge and attitudes of practitioners in the field and legislators as to the relationship between test knowledge and attitude and instructional decision making. He investigated the level of knowledge of tests that educators and legislators possessed, how test data were used in instructional decision making, and the attitudes of teachers and legislators toward using tests. To determine the level of knowledge about tests, for example, individuals were shown a sample of reading test scores from the California Assessment Program for two separate years and asked to interpret the scores and the differences between them. They were also asked if they or their colleagues could understand certain kinds of test scores and how they would rank them for use. Surprisingly, most teachers, 85%, indicated they would have difficulty understanding some scores. Among those, were the total test score, and, as part of a

larger test, raw scores for sub tests, percentile scores for sub tests, percent correct for sub tests, and grade scores for sub tests. Many of the teachers said they would require some assistance interpreting these scores.

To determine how test data were used and what effect tests had on instructional decisions, Ruddell (1985) used a preestablished interview protocol. All groups were asked to discuss their use of test data and how it affected their educational decisions. Most responded that test data were not the major factor in making decisions. The most common use of test data was by teachers and principals in assessing progress of the school reading programs. The legislators agreed that they had little use for the data. One admitted that the comprehensive achievement tests administered state-wide "should be used by administrators to evaluate and improve instruction," but he did not believe it was being done (p. 540).

Interview discussions about general testing philosophy and test use examined attitudes toward testing and using tests to make decisions about students. Not surprisingly, according to Ruddell, as the distance from the classroom increased, the use of tests changed. Teachers, closest to the students, were more in favor of tests that were diagnostic in nature. Principals worried more about the match of the test to the curriculum and for program evaluation. Central office personnel were more attentive to trends and state-mandated testing. Legislators were not interested in test scores, per se.

Ruddell's study recognized many of the fallacies that surround standardized achievement testing. He found, for example, that the knowledge of teachers, administrators, and legislators about basic test ingredients was very limited, contrary to what most education "experts" would have us believe. Also, he found that standardized test results were not used very much in educational decisions. More use was made of test scores to assess programs and identify trends. Teachers tended to rely more on diagnostic tests and personal observation to assist students with their educational needs.

One prevalent attitude felt by most respondents in his study was that testing took too much time away from instruction.

Brown (1989), concerned with the new emphasis on both informational and test literacy, examined the role of testing in determining educational policies in curriculum development. Almost everyone he interviewed determined educational success or progress on the basis of scores on standardized, norm-referenced tests. Also, in keeping with the national trend, he found that those who had more at stake based on the results of the test attached more importance to it. Those schools with no pressure made little use of the results of the testing; those where funding or other important aspects were tied to achievement test scores tended to "teach more to the test" than not. Additionally, more significance was attached to the rise and fall of test scores in those districts. In keeping with other research, Brown also found that most people he talked to believed "that commercial standardized tests were valid indicators of student and school achievement" (p. 31).

While concerned more with teacher-made tests, Marso and Pigge (1989) examined testing practices, attitudes, and test construction proficiencies among teachers. Their results led to some generalizations about knowledge and attitudes toward standardized tests as well. Marso and Pigge surveyed two groups. The first was 800 Ohio public school supervisors and principals. The second was 600 Ohio public school teachers, all of whom had been required to take a tests and measurement course during their teacher training. In addition, all the teachers were asked to submit a copy of their most recently developed formal teacher-made test along with their completed survey instrument. "Both the teachers and the administrators reported a high need for teachers to possess skills in the use of teacher-made and standardized tests in day-to-day instruction to assure teachers's success in the classroom" (p. 7). A consensus regarding their perceptions appeared to be reached among teachers, principals, and supervisors

that testing skills are among teachers' weaker professional proficiencies. Additionally, Marso and Pigge suggested that "testing specialists and/or practicing educators need to reexamine the role and value of test construction statistics for classroom teachers" (p. 8). They also found that many universities do not offer a separate course in tests and measurement in their teacher training programs.

Stiggins (1991) defined what he termed "assessment literacy." He suggested that those who are "assessment literate" have a basic understanding of not only what it takes to produce high-quality achievement data, but also the ability to critically evaluate what data they do have. It is not enough to be able to recognize high-quality data from low-quality data, proposed Stiggins, but one must also recognize extraneous factors that affect those data. Also concerned with assessment literacy, Johns and Davis (1991) and VanLiersburg and Johns (1991) examined preservice and inservice teachers as to their perceptions of standardized testing procedures, actual test-taking, and ethical considerations associated with both test-taking and test administration. In both studies, a diversity of opinions about what the term "standardized" meant was found. In fact, their results showed that

large numbers of both groups did not fully understand the importance of standardized test procedures and that over one third of both preservice and inservice teachers were unclear about the ethics of test-wiseness programs as contrasted to actual teaching to the test (VanLeirsburg & Johns, 1991, p. 5).

Both studies suggested that preservice as well as inservice teachers could benefit by possessing more knowledge about standardized tests and testing procedure (Johns & Davis, 1991; VanLeirsburg & Johns, 1991).

Marso and Pigge (1992) reviewed 225 studies on teachers' knowledge of tests and identified several points of concern. Among them were:

1. Many practicing teachers report having received no formal measurement training during their preservice education, many practicing teachers report having received just a single unit of instruction in measurement within another preservice education course, and most practicing teachers report

having received no school-sponsored inservice training in the development and use of classroom tests.

2. Teachers with typical formal training in tests and measurement perform better than teachers without this training on paper and pencil measures of testing knowledge, but their scores typically exceed the scores of untrained teachers by just six to ten percent.
3. Teachers have difficulty in correctly answering questions related to appropriate interpretations of scores commonly used in conveying pupil performance on standardized and state competency tests.
4. Teachers display especially limited knowledge about technical aspects of testing (e.g., use of test specification tables, item analysis and statistical analysis procedures, etc.)
5. Teachers, principals, and supervisors appear to agree rather highly one with another about the relative level of teachers' proficiencies in various testing skills; they also agree one with another that teachers' preservice preparation in testing is less adequate than their level of preparation in other areas of professional training (Marso & Pigge, 1992, pp. 26-28).

From the literature on teachers' knowledge about tests and testing, it appears that the level of knowledge of current practitioners is not more than it was when Roeder made his suggestion in 1973. Twenty years have passed, and it seems that teachers, principals, administrators, colleges of education, and legislators have not increased their level of knowledge in keeping with the information explosion and advocacy for test use that has engulfed the country. It is hoped that work such as this will lead to increased professional preparation of teachers and administrators.

CHAPTER III

METHODOLOGY

The Population

The population for this study consisted of a randomly-selected sample of 400 teachers and administrators in the state of Tennessee. The sample population was selected from a master list of all educators in the state maintained by the Tennessee Education Association (TEA). A random-ordered list of 400 names and addresses of all educators currently employed in the state, not just TEA members, was provided by the local TEA representative.

Data Collection Procedures

Data were obtained from the sample by means of a mailed survey instrument, a copy of which may be found in Appendix A. This survey instrument was the final step of a three-step process which utilized both a preliminary survey and a pilot survey. The first step, the preliminary survey, employed an early form of the final instrument that was hand-administered to personally-known educators. Suggestions were solicited as to physical arrangement, wording, clarity, ambiguity, or other impediments to understanding. The answers to the preliminary survey questions were not used in the final data analysis. After reviewing those educators' suggestions and making modifications, a pilot survey form was devised.

The second step utilized a modified form of the preliminary survey instrument constructed from the preliminary survey instrument. The population for the pilot study consisted of a sample of 100 individuals randomly selected from the teachers and administrators in Roane County, Tennessee. A letter granting approval to use the

educators in Roane County for this purpose, a copy of which may be found in Appendix B, was obtained from the Superintendent of Roane County Schools. A listing of the currently employed certified teachers and administrators was obtained. A cross-check of personnel rehired for the 1993-94 school year was made to validate the current staff. Those who had resigned or retired were deleted; non-certified personnel such as maintenance or secretarial personnel were eliminated, leaving a total of 343 personnel. Using a table of random numbers, 100 names were selected. This population was mailed a cover letter stating the reason and the method of the survey, a copy of the pilot questionnaire, and a self-addressed, stamped envelope. The purpose of this pilot survey, like the preliminary survey, was to determine if the survey questions would provide the needed information, if the questions were appropriate for the final population to be surveyed, and if the information received was consistent with the stated purposes of the study. Of the 100 pilot survey questionnaires mailed, 42% were returned. The data obtained from the preliminary survey were not used in the final analysis of data. Recommendations from the respondents of the preliminary survey were used to prepare the final instrument. Several questions were combined into one, and four questions were eliminated leaving a total of 26 items on the final instrument.

The third step of the study utilized the final survey instrument devised as a result of the pilot survey. The items on the final questionnaire were classified into three broad categories: 1) use of standardized test results, 2) attitudes toward standardized achievements tests used as a component of teacher evaluation, and 3) the preparation and perception of respondents to properly identify and interpret standardized test results. The survey questions were arranged such that items in each category tended to be grouped together.

The final questionnaire was mailed to the sample in October of 1993. Enclosed in the mailing were a cover letter, Appendix C, explaining the purpose of the survey, the final survey instrument, and an addressed, postage-paid return envelope. The return

envelope was coded with the number appearing on the address label for determination of the respondent; the survey instrument was not coded in any way. Questionnaires and return envelopes were separated at opening to ensure the anonymity of the respondents.

Of the 400 questionnaires mailed, 101 were returned for an initial return rate of 25.25%. Due to the low rate of return, a second mailing was made in the spring of 1994 to those members of the sample who had not responded to the initial mailing. This second mailing, like the first, consisted of a cover letter, Appendix D, explaining the purpose of the survey, the final survey instrument, and an addressed, postage-paid return envelope coded as before. Approximately two weeks after the second mailing, postcards were mailed as a reminder. Approximately three weeks after the postcard, a third mailing was sent. The third mailing also consisted of a cover letter similar to the one contained in the second mailing, a questionnaire, and an addressed, postage-paid return envelope.

The response to the second and third mailings was an additional 126 responses for a total return of 227, a rate of 56.75%. Ten of those questionnaires were returned blank, with various comments such as "retired," "don't use tests in P.E.," or "I teach Special Ed. and don't use standardized tests," an unusual reply because students are assigned to special education based on the results of testing. The remainder was 217 usable instruments from a total population of 390 for a return rate of 55.64%. Of the remaining 173 non-respondents, listed telephone numbers were obtained for 128. Over a three-day period, 75 of that number were contacted by telephone. Those contacted over the telephone were asked for demographic data only--age, years' experience, grade level taught, highest educational degree, and current position--to determine if they matched the profile of the respondents who had provided written data. Five subsequently returned the questionnaire, 4 subjects declined to participate in the survey and 1 subject was deceased. The total number of returns was 222 for a response rate of 56.92% from a total population of 390.

Analysis of the Data

Statistical analyses of the data in the study were performed using the Statistical Package for the Social Sciences (SPSS), Release 4.1, currently running on the VAX/VMS cluster at the University of Tennessee, Knoxville. This computer program offered many and varied statistical procedures that were available for the analysis of data including those used in this study: frequency analysis, descriptive statistics--mean, standard deviation, minimum and maximum scores, and number--and cross tabulations between variables.

A frequency analysis was performed on the demographic portion of the data, i.e. age, years' experience in education, current grade level, highest educational attainment, work on an advance degree, and current assignment. The rationale for this analysis was to determine a profile of the respondents to the survey. It was believed that older, more experienced educators who had earned advanced degrees would have a greater knowledge of testing and test use than younger, less experienced personnel. Throughout their teaching careers, many teachers are exposed to testing by virtue of their position in the classroom. Elementary teachers tend to have more experience with testing and its results than secondary teachers as elementary students are tested more for various reasons, e.g., placement in special programs. As years' experience increases, more and more exposure occurs leading to a greater knowledge of test procedures and test use. Also, professional educators tend to increase their professional knowledge by attending classes and working toward advancement in their field. This, too, expands their knowledge about tests and measurement as they participate in various college courses aimed at understanding psychological measurement. It was thought that educators who had attained advanced degrees would have a greater knowledge of testing and terms associated with tests and measurement. As educators gain more experience, they come to rely more on tests, both standardized and not, to make some of the important decisions

concerning their students. Consequently, it was likely that educators who had greater years of experience would have more knowledge about standardized testing and its terms and would use that knowledge to assist them in making educational decisions about their students. Also, educators newer to the profession could be expected to have benefited from better teacher preparation in the area of tests and measurement.

To enable the answering of the research questions, frequency analyses were performed on the following items (applicable research questions appear in parentheses):

1. the context in which testing terms were used (2),
2. whether or not the respondent had taken a course in tests and measurement (3),
3. a description of the course if one had been taken (3),
4. how frequently material from that course was utilized in respondent's current work (3),
5. how frequently respondents had undertaken work on their own to increase their knowledge of tests and measurement (3),
6. utilization of the technical manual that accompanies standardized tests (3),
7. the frequency with which schools/districts offer activities to increase teachers' knowledge of tests and measurement (3),
8. experience of respondents with regard to teacher evaluation based on standardized test scores (4), and
9. response to subsequent teacher evaluation using test scores (4).

Cross tabulations were calculated between the demographic data and certain other variables to determine relationships among them. These cross tabulations included:

1. age by use of standardized test scores (1),
2. age by attitudes toward standardized tests (4),
3. age by attitudes of using tests for teacher evaluation (4),

4. age by perception of term knowledge (2),
5. years' experience by use of standardized test scores (1),
6. years' experience by educational decisions (1),
7. years' experience by attitudes toward standardized tests (4),
8. years' experience by attitudes of using tests for teacher evaluation (4),
9. years' experience by perception of of term knowledge (2),
10. current assignment by use of test scores (1),
11. working on advanced degree by age (2),
12. working on advanced degree by years' experience (2), and
13. working on advanced degree by educational level (2).

The rationale underlying these cross tabulations was that the demographic variables, age, years' experience, and so forth, would have had an effect on the respondents' answers to the items on the survey. For example, it was thought that older respondents, those with more years' experience, and those who had earned a higher level degree would possess more knowledge about standardized testing and use those terms and concepts more readily than those who were younger, less experienced, or who had not returned for additional college course work. Consequently, cross tabulations comparing these variables with demographic variables provided such information.

Also, cross tabulations were also performed between other variables to determine if relationships existed among them. Those additional cross tabulations included:

1. context of use of testing terms by use of testing terms (1),
2. courses taken in tests and measurement by utilization of course material (3),
and
3. courses taken in tests and measurement by use of technical manual (3).

The rationale underlying these cross tabulations dealt with the degree of use of material learned in tests and measurement courses and the use of the technical manual

that accompanies standardized tests. It was thought that those who had taken such courses would be expected to utilize that material more if they perceived that it had some value. Responses to these items showed the value the respondents placed upon the information received in that course or courses.

This study attempted to determine profiles of the respondents' attitudes towards standardized testing and the use of students' test data as a component of teacher evaluation. As such, the resultant data has been reported in frequencies and percentages.

CHAPTER IV

PRESENTATION AND ANALYSIS OF THE DATA

Population Demographics

A profile of the population of the study was determined by six demographic questions. Respondents were asked to check categories for their age, years' experience in education, the grades they currently taught, their highest educational level, whether or not they were currently working on an advanced degree, and their current assignment. Those who were contacted by telephone provided the answers to five questions, omitting the question concerning current work on an advanced degree. These parameters have a relationship with the other items in the survey in that other responses will be affected by them. For example, it was likely that a person with an advanced degree would have more knowledge about testing and standardized tests than one who did not. Or, one who had more years' experience than another would have had more exposure to standardized testing and would thus be more knowledgeable about testing terms and processes.

On the written survey, respondents were requested to identify their age by marking 1 of 6 increments of 5 years' range beginning with 20-25 years old and ending with a 7th response of "greater than 50." According to the replies from both the survey instrument and phone contacts, the majority of the respondents, 25.2%, were between the ages of 41-45. The next largest group, 24.8%, were greater than the age of 50. The 3rd largest group, 16.7%, were between the ages of 46 and 50. As a result, two-thirds of the respondents to the survey indicated they were greater than the age of 41. See Table 1.

Second, the respondents were asked their years' experience in education. The

Table 1

Age of Respondents

Item	Responses	
	Percent ^a	N
1. 20 - 25 years	1.8	4
2. 26 - 30 years	13.1	29
3. 31 - 35 years	6.3	14
4. 36 - 40 years	12.2	27
5. 41 - 45 years	25.2	56
6. 46 - 50 years	16.7	37
7. Greater than 50 years	24.8	55
8. Missing Responses	0.0	0
Total		222

^aAdds to more than 100% due to rounding

years of experience of this group tended to follow the same pattern as the age demographic. The largest percentage, 19.8%, reported they had had between 21 and 25 years' experience in education. The next largest group, 16.7%, stated they had between 5 and 10 years' experience. The least percentage, 9%, had less than 5 years' experience. The years' experience tended to be weighted toward the higher end similar to the age demographic, with the majority, 57.7%, having greater than 16 years' experience in education. See Table 2.

Just as the age and years' experience of the respondents were important, so were the grades taught by them. Standardized tests tend to be used more in elementary grades than in secondary grades. Students tend to be classified more in earlier years because of the desire to place them in the correct educational setting. Elementary level educators, therefore, tend to be more familiar with tests and measurement than secondary personnel. Thus, the third item on the survey asked for the current grade level of the respondents. The responses were listed from Early Education (K-3) to a System-wide position. The responses are listed in Table 3.

The responses to this item showed more elementary and middle school teachers responding to the survey than did secondary teachers and system-wide personnel. The largest individual percentage of respondents, 25.7%, identified themselves as elementary teachers, Kindergarten to 6th grade. The next largest percentage, 23.9%, also identified themselves as elementary school teachers in the early elementary grades, Kindergarten through 3rd grade. The remainder, 46.9%, either identified themselves as at the middle school, junior high, secondary level, or system-wide; 3.6% of the respondents did not identify a grade level.

Fourth, the educational level of the respondents was asked. As stated earlier, it was thought that a distinct relationship existed between a person's level of education and his knowledge about standardized tests. Because most advanced degrees in education require some study in the area of tests and measurement, it was believed that those who

Table 2

Years' Experience in Education

Item	Responses	
	Percent ^a	N
1. Less than 5 years	9.0	20
2. 5 - 10 years	16.7	37
3. 11 - 15 years	15.8	35
4. 16 - 20 years	14.9	33
5. 21 - 25 years	19.8	44
6. 26 - 30 years	12.6	28
7. Greater than 30 years	10.4	23
8. Missing responses	.9	2
Total		222

^aAdds to more than 100% due to rounding

Table 3

Grades Currently Taught

Item	Responses	
	Percent ^a	N
1. Early Elementary (K-3)	23.9	53
2. Elementary (K-6)	25.7	57
3. Middle School (6-8)	13.5	30
4. Junior High (7-9)	7.2	16
5. High School (9-12)	13.5	30
6. Senior High (10-12)	7.7	17
7. System-Wide	5.0	11
8. Missing responses	3.6	8
Total		222

^aAdds to more than 100% due to rounding

had completed an advanced degree would have more knowledge in this area. The respondents were asked the highest educational level they had attained, from a Bachelor's degree (BS or BA) to a Doctoral Degree (either Ph.D. or Ed.D.). A response of "Other" was allowed, with 1.4% choosing that response. The majority of the respondents, 40.1%, reported they had either a BS or BA degree. The next largest group, 33.8%, said they had attained a Master's degree, either a Master of Science (MS) or a Master of Arts (MA). An educational level of a Master's plus 45 hours beyond the Master's was reported by 17.1%. Only 4.5% reported an Ed.S. degree, and 1.8% reported a doctoral degree. Of the 1.4% who reported "Other," 1 was a certified vocational teacher, 1 reported a BS + 15 hours, and 1 reported a Master of Fine Arts (MFA) degree. Three did not respond to this item. See Table 4.

Fifth, in addition to the highest educational level attained, respondents were also asked if they were currently working on an advanced degree. Such work might indicate a more than passing knowledge of standardized testing and its terms for the same reasons stated above. Of the respondents, 10.8% stated that they were currently working on an advanced degree, while a large majority, 82%, stated they were not. An undecided response was given by 5.9%, and 1.4% did not respond to the question.

A comparison was made between those respondents who indicated they were working on an advanced degree and their age, years' experience, and educational level. The results are shown in Table 5. By age, the largest percentage of groups who indicated work on an advanced degree were in the 26 to 30 age group and the 20 to 25 age group. As educators begin their professional careers, it is not unusual for them to work on an advanced degree; therefore, the younger age group represented here indicated that. Concurrently, by years' experience, the largest percentage of groups working on an advanced degree were those with between 0 and 10 years' experience. As might be expected, the largest percentage of those groups working on an advanced degree were those with a bachelor's degree and a master's, 28% total. Yet a fairly large number,

Table 4

Highest Educational Level

Item	Responses	
	Percent ^a	N
1. BS/BA	40.1	89
2. MS/MA	33.8	75
3. MS/MA + 45 Hours	17.1	38
4. Ed. S.	4.5	10
5. Ph.D./Ed.D.	1.8	4
6. Other	1.4	3
7. Missing responses	1.4	3
Total		222

^aAdds to more than 100% due to rounding

Table 5

Working on Advanced Degree by Age, Years' Experience, and Educational Level

Items	Responses ^a						Total N
	%Yes	N	%No	N	%Undecided	N	
1.Age:							
20 - 25	25.0	1	75.0	3	0.0	0	4
26 - 30	31.0	9	62.1	18	6.9	2	29
31 - 35	7.1	1	85.7	12	7.1	1	14
36 - 40	4.0	1	84.0	21	12.0	3	25
41 - 45	14.3	8	83.9	47	1.8	1	56
46 - 50	10.8	4	78.4	29	10.8	4	37
> 50	0.0	0	96.3	52	3.7	2	54
Missing							3
Total							222
2. Years' Experience:							
< 5	25.0	5	75.0	15	0.0	0	20
5 - 10	22.2	8	66.7	24	11.1	4	36
11 - 15	11.8	4	82.4	28	5.9	2	34
16 - 20	9.1	3	84.8	28	6.1	2	33
21 - 25	6.8	3	86.4	38	6.8	3	44
26 - 30	3.6	1	89.3	25	7.1	2	28
> 30	0.0	0	100.0	23	0.0	0	23
Missing							7
Total							222
3. Educational Level:							
BS/BA	17.2	15	77.0	67	5.7	5	87
MS/MA	10.8	8	85.1	63	4.1	3	74
MS/MA + 45	2.6	1	89.5	34	7.9	3	38
Ed. S.	0.0	0	90.0	9	10.0	1	10
Ph. D./Ed/ D.	0.0	0	100.0	4	0.0	0	4
Other	0.0	0	100.0	3	0.0	0	3
Missing							6
Total							222

^aAdds to more than 100% due to rounding

77% of those with only a bachelor's and 85.1% of those with a master's degree indicated that they were not pursuing an advanced degree.

The majority of the respondents, 81.5%, stated their current assignment was as a classroom teacher; 5.4% were a principal or assistant; 3.6% were a system-wide administrator; and 8.1% indicated they were in some other assignment, either a health and physical education instructor shared among different schools, a traveling reading teacher, a music teacher, or a special education teacher.

The respondents who were interviewed by telephone gave similar data. The comparison is shown in Table 6. The greatest differences in the age category was in the 26 to 30 age group, 8.4 percentage points, and in the 36 to 40 age group, 8.1 percentage points. Both these differences are below a 10% level. In the Years' Experience category, the largest difference was in the group with between 11 and 15 years' experience. Here the difference was 10.8 percentage points, only marginally above a 10% level. Another difference existed in this group with those between 5 and 10 years' experience. The difference here was 8.9%, also below 10%. In the Grades Taught category, those who reported teaching in the elementary grades on the survey instrument was 49.6%; in the telephone survey, those who indicated they taught in the elementary grades was 50%, a very close match. The largest discrepancy in this area was between those who reported they currently taught at the high school level. The difference between the survey respondents and the telephone respondents was 14.6 percentage points. In the category of Educational level, the largest difference existed in the number of respondents who indicated they had a MS/MA. Of those in the survey, 33.8% indicated they had earned a MS/MA degree while 48.4% contacted by telephone indicated they had earned an MS/MA degree, a difference of 14.6 percentage points. In the category of Current Assignment, no principals were contacted by phone, and only one System-wide administrator was contacted. Based on the comparison of the respondents' demographic data, it can be reasonably assumed that those who were contacted by

Table 6

A Comparison of Survey Respondents and Telephone Respondents by Age, Years'Experience, Grades Taught, and Educational Level

Items	Responses ^a				
	%Survey	N	%Telephone	N	Total N
1. Age:					
20 - 25	1.8	4	0.0	0	4
26 - 30	13.1	29	4.7	3	32
31 - 35	6.3	14	9.4	6	20
36 - 40	12.2	27	20.3	13	40
41 - 45	25.2	56	23.4	15	71
46 - 50	16.7	37	15.6	10	47
> 50	24.8	55	26.6	17	72
Total		222		64	286
2. Years' Experience:					
< 5	9.0	20	4.7	3	23
5 - 10	16.7	37	7.8	5	42
11 - 15	15.8	35	26.6	17	52
16 - 20	14.9	33	17.2	11	44
21 - 25	19.8	44	18.8	12	56
26 - 30	12.6	28	14.1	9	37
> 30	10.4	23	10.9	7	30
Missing	0.9	2	0.0	0	2
Total		222		64	286
3. Grades Taught:					
Early Elementary School	23.9	53	10.9	7	60
Elementary School	25.7	57	39.1	25	82
Middle School	13.5	30	12.5	8	38
Junior High School	7.2	16	0.0	0	16
High School	13.5	30	28.1	18	48
Senior High School	7.7	17	3.1	2	19
System-Wide	5.0	11	3.1	2	13
Missing	3.6	8	3.1	2	10
Total		222		64	286

Table 6 (continued)

Items	Responses ^a				
	%Survey	N	%Telephone	N	Total N
4. Educational Level:					
BS/BA	40.1	89	37.5	24	113
MS/MA	33.8	75	48.4	31	106
MS/MA + 45	17.1	38	4.7	3	41
Ed. S.	4.5	10	4.7	3	13
Ph. D./Ed/ D.	1.8	4	0.0	0	4
Other	1.4	3	4.7	3	6
Missing	1.4	3	0.0	0	3
Total		222		64	286
5. Current Assignment:					
Classroom teacher	81.5	181	87.5	56	237
Principal	5.4	12	0.0	0	12
System-wide administrator	3.6	8	1.5	1	9
Other	8.1	18	10.9	7	25
Missing	1.4	3	0.0	0	3
Total		222		64	286

^aAdds to more than 100% due to rounding

telephone would have indicated much the same responses as those who returned the survey questionnaire.

Due to the method of data collection, one in which the initial survey instrument was distributed in the fall and the follow-up instrument distributed in the spring, there existed a possibility that respondents might have answered differently based on the time lapse between the two distributions. A comparison was made between the frequencies of responses of the two groups to certain key items in the survey questionnaire. The key items chosen to compare were those directly associated with the research questions that guided the study, i.e. Test Score Use, Attitudes toward Standardized Tests, and Attitudes toward Value-added Assessment, Term Perception, and Prior Training. That comparison is shown in Table 7. The largest difference between means appeared in the item dealing with Attitudes toward Standardized Tests. A difference of .48 existed between the fall mean and the spring mean for the section that asked about testing for lower-order thinking skills. Yet, within the same item, there was no difference between the means for the section that dealt with testing higher-order thinking skills. The next highest difference was in the item that dealt with Test Score Use. A difference of .43 appeared in the section that asked about the use of standardized tests for policy formation. Within that same item a difference of .41 appeared in the section that asked if standardized test scores were used to report to the district. No explanation for those three differences exists. All other items showed minor or negligible differences in means between the fall survey and the spring survey. Based on this comparison of the responses to these items, it was reasonably assumed that the time lapse between the mailing of the two survey instruments would have had no effect on the responses from either set of respondents. That is, the respondents who returned the spring questionnaire answered approximately the same as the ones who returned the fall survey instrument.

The profile of the respondents to this survey, therefore, denoted that of an elementary school classroom teacher with a BS or BA degree not currently working on an

Table 7

A Comparison of Fall Respondents and Spring Respondents by Test Score Use, Attitudes toward Standardized Tests, Attitudes toward Value-Added Assessment, Term Perception, and Prior Training

Items	Responses		
	Fall Mean	Spring Mean	Difference Fall-Spring
1. Test Score Use			
Allocation of funds	1.15	1.33	-.18
Assigning students to classes/programs	3.16	3.32	-.16
Communicating to parents	3.54	3.28	.26
Curriculum planning	3.31	3.17	.14
Diagnosing skill problems	3.63	3.61	.02
Informing the public	3.08	2.94	.14
Policy formation	2.22	1.79	.43
Reporting to district	3.02	2.61	.41
Student Promotion/retention	2.37	2.21	.16
Teacher evaluation	2.19	2.21	-.02
2. Attitudes toward Standardized Tests			
Reliable means to assess students	2.46	2.35	.11
Valid means to assess students	2.43	2.41	.02
Reliable method for diagnosing students	2.55	2.46	.09
Test the cognitive domain	3.13	2.93	.20
Test the affective domain	2.16	2.27	-.11
Test higher-order thinking skills	2.56	2.56	0.00
Test lower-order thinking skills	3.25	2.77	.48
3. Attitudes toward Value-added Assessment			
Should be component of teacher evaluation	1.73	1.63	.10
Valid means of assessing teachers	1.47	1.55	-.08
Reliable means of assessing teachers	1.46	1.52	-.06
Indicate teacher effects on students	2.07	1.89	.18
Accurate picture of student's learning	1.67	1.75	-.08
Reflect student's achievement gains	2.09	2.26	-.17

Table 7 (continued)

Items	Responses		
	Fall Mean	Spring Mean	Difference Fall-Spring
4. Term Perception			
Adjusted score	2.55	2.61	-.06
Criterion-referenced test	3.52	3.30	.22
Grade equivalents (GE)	3.93	3.89	.04
Mean	4.14	4.09	.05
Median	4.07	4.06	.01
Mode	3.73	3.85	-.07
Normal distribution	3.69	3.63	.06
Norm-referenced test	3.87	3.77	.10
Percentile	4.35	4.26	.09
Percentile rank	4.36	4.17	.19
Raw score	4.03	3.98	.05
Reliability	3.40	3.46	-.06
Scale Score	3.22	3.20	.02
Standard Score	3.58	3.49	.09
Stanine	3.98	3.78	.20
Statistical process control	2.22	2.10	.12
Validity	3.25	3.30	-.05
5. Prior Training	N	N	Total
YES, complete course	54	67	121
YES, partial course	21	22	43
NO	13	25	38
DON'T REMEMBER	5	5	10
Missing			10
Total			222

advanced degree, greater than 41 years of age with 21 to 25 years of experience in education.

Data Analysis

Use of Standardized Test Score Data

The first question the study posed dealt with the use of student standardized test data by teachers and administrators in the state of Tennessee. Respondents were asked about their use of student standardized test scores based on their current assignment, as a classroom teacher, a principal, or a system-wide administrator. The responses--with no allowance for assignment--and the number responding are shown in Table 8.

As can be seen, the most prevalent use of test score data by practitioners in Tennessee was for diagnosing skill problems among students. This stated use was in keeping with a frequently-stated use of standardized test scores as identified by other studies, but was not considered to be the most prevalent use of test data (Linden & Linden, 1968; Madaus, 1991). That use was designated to the utilization of standardized test data for a planning function. The next most common use according to this survey was the use of test score results to assign students to classes or programs. As far as the respondents of the survey were concerned, the most infrequent use of test scores was for the allocation of funds, perhaps because teachers do not have much say-so, if any, in how school or district funds are to be allocated. On that topic, 36.2% of the teachers indicated that they did not know how often standardized test scores were used for the allocation of funds; 39.1% indicated that test scores were never used for that purpose. By contrast, only 5.7% of teachers stated that test scores were frequently used for the allocation of funds within their classroom. Of the 11 principals who responded to this item, 81.9% indicated that standardized test scores were never or sometimes used for the allocation of funds in their schools, and much the same result was obtained from the

Table 8

Overall Use of Test Score Results

Item	Responses	
	N ^a	Mean ^b
Diagnosing skill problems	218	3.71
Assigning students to classes/programs	219	3.49
Communicating to parents	218	3.46
Reporting to district	216	3.45
Curriculum planning	218	3.44
Informing the public	217	3.34
Policy formation	217	2.76
Teacher evaluation	219	2.62
Student promotion/retention	219	2.43
Allocation of funds	213	1.94

^aNumber responding (of the 222 participants)

^bResponses were from 1= Never to 5 = Frequently with 3 = Sometimes

eight system-wide administrators who responded, with 62.5% responding that same way. Consequently, it would appear that standardized test scores are not often used to determine the allocation of funds within classrooms, schools, or districts.

When asked about the assignment of students to classes or programs, 71% of the teachers responded that standardized test scores were sometimes or frequently used for that purpose. All the principals and seven of eight administrators gave the same response. Over 80% of the teachers gave reporting to parents as one of the major uses of standardized scores, with 31.5% stating a sometimes use, 22.5% a more than sometimes use and 24.7% giving a frequent use. Of the principals and administrators who responded, again, all placed the use of test scores to report to parents at or above the sometimes mark. For curriculum planning, 75.2% of the teachers said it was sometimes to frequently used in their classrooms, and all of the principals responded likewise. Only one administrator stated that test scores were not used that much within his district for curriculum planning, ranking it just below sometimes. Almost four-fifths, 79.2%, of the teachers stated they sometimes to frequently used test scores for diagnosing skill problems, with 34.3% stating it was a frequent use thus reinforcing the earlier findings that this was the most prevalent use of standardized test scores among the respondents. All of the principals and 87.5% of the administrators marked this item as a sometimes or frequent use of test scores. Again, one administrator did not see this as a major use for test scores in his district, scoring it as less than sometimes.

As far as using standardized test scores to inform the public, 62.2% of the teachers indicated that this was a sometimes or frequent use of scores, with 27.7% stating it was a frequent use. The principals were not so certain, with only 41.7% rating it sometimes and 25.0% rating it less than that. Four principals saw it as above the median mark, but not as a frequent use. The administrators were firmly in the middle with 50% ranking it as sometimes. As an impetus for policy formation, most of the teachers did not see test scores as being used for that purpose, with 49.1% ranking

it sometimes or below and 28.2% indicating they did not know. Only 11.9% indicated that test scores were frequently used for policy formation. The principals agreed, with half stating that standardized test scores were infrequently used for policy formation in their schools and, 33.3% indicating they did not know if test scores were used for that purpose. The administrators, dealing more with policy formation at their level, stated that standardized test scores played a larger role in policy formation with three-fourths indicating a score of sometimes or above. When the use involved reporting to the district office, almost 56.8% of the teachers indicated that this was a sometimes to frequent use of test scores, with the majority, 29.5%, indicating a frequent use. Almost all the principals saw this as an important use, with 74.9% placing it at sometimes or above. The administrators ranked this item lower with three-fourths giving it a low to sometimes ranking and the other quarter rating it frequently.

In the area of student promotion or retention, the majority of the teachers, 76%, said that test scores were never or sometimes used for this purpose. By contrast, the principals and administrators were equally split in their view of the use of test scores for student promotion or retention. Half of the principals said that test scores were sometimes or frequently used for promoting or retaining students, and half of the administrators rated that use at sometimes or above. This indicated a division between the view of the use of test scores for this purpose between teachers and administrators. It could be that teachers were viewing this use in their classes only, while principals and system-wide administrators were looking at the overall school and district respectively; yet this did suggest that teachers and administrators viewed the use of test scores for the promotion and retention of students in a different way.

When the use of test scores for teacher evaluation was asked, 58.5% of the teachers said that test scores should never or sometimes be used. Only 15.9% indicated that they did not know, and only 14.2% indicated that they should frequently be used for teacher evaluation. All but one of the principals ranked the use of test scores for teacher

evaluation at the low end of the scale, from never to sometimes with one principal indicating one rank above sometimes. The administrators, too, put this use at the low end, with 25.0% of the administrators stating they did not know if standardized test scores should be used for this purpose, 25.0% ranking it never, with only one ranking it above sometimes. The rankings and percentages are displayed in Table 9.

When the age of the respondents was matched with the score use, much the same picture emerged. The results of that comparison are listed in Table 10. Of those greater than 50 years old, 51.9% stated that standardized test scores were never used for allocation of funds. Of those between 46 and 50 years old, 33.3% stated the same; 43.6% between 41 and 45 years old gave the same response. In all cases, the majority in each age division who provided a response between never and frequently, stated that standardized test scores were never used for fund allocation. When asked about assigning students to classes or programs, all respondents greater than 31 years of age stated that this function was done frequently, with the only difference appearing in the 46 to 50 age group where the majority response was sometimes. As the age of the respondent decreased, so did the scores for the use of assigning students to classes or programs.

When asked about using test scores to communicate to parents, the pendulum swung the other way. Matching the earlier findings, the majority in almost every age group responded that test scores were sometimes or frequently used to communicate to parents, with the largest numbers found in the older age groups. As the age increased, the likelihood of using test scores to report to the parents increased.

In the area of curriculum planning, there was a match between the older and younger respondents. Both groups indicated that test scores were sometimes to frequently used for curriculum planning with the majority in the middle age group indicating that this was a sometimes use. Among the younger than 36 and older than 50 age groups, the majority indicated it as a frequent use. For diagnosing skill problems, the older respondents, greater than 50, indicated that this was a frequent use of test

Table 9

Use of Test Scores by Assignment

Item	Percentage ^a of Responses						N
	0 ^b	1	2	3	4	5	
1. Allocation of funds							
Teachers	36.2	39.1	8.6	6.3	4.0	5.7	174
Principals	18.2	45.5	18.2	9.1	9.1	0.0	11
Administrators	25.0	12.5	37.5	12.5	12.5	0.0	8
Other	38.9	16.7	5.6	27.8	11.1	0.0	18
Missing							11
Total	35.1	36.5	10.0	8.5	5.2	4.7	222
2. Assigning students to classes/programs							
Teachers	6.7	12.3	10.1	28.5	11.2	31.3	179
Principals	0.0	0.0	0.0	33.3	33.3	33.3	12
Administrators	0.0	0.0	12.5	25.0	12.5	50.0	8
Other	11.1	16.7	0.0	16.7	22.2	33.3	18
Missing							5
Total	6.5	11.5	8.8	27.6	13.4	32.3	222
3. Communicating to parents							
Teachers	1.7	11.8	7.9	31.5	22.5	24.7	178
Principals	0.0	0.0	0.0	16.7	58.3	25.0	12
Administrators	0.0	0.0	0.0	37.5	37.5	25.0	8
Other	5.6	5.6	5.6	50.0	11.1	22.2	18
Missing							6
Total	1.9	10.2	6.9	32.4	24.1	24.5	222
4. Curriculum planning							
Teachers	5.1	10.7	9.0	33.7	20.2	21.3	178
Principals	0.0	0.0	0.0	25.0	58.3	16.7	12
Administrators	0.0	0.0	12.5	0.0	50.0	37.5	8
Other	22.2	5.6	0.0	22.2	33.3	16.7	18
Missing							6
Total	6.0	9.3	7.9	31.0	24.5	21.3	222

Table 9 (continued)

	Percentage ^a of Responses						
Item	0 ^b	1	2	3	4	5	N
5. Diagnosing skill problems							
Teachers	2.8	6.2	11.8	21.9	23.0	34.3	178
Principals	0.0	0.0	0.0	41.7	50.0	8.3	12
Administrators	0.0	0.0	12.5	12.5	25.0	50.0	8
Other	0.0	5.6	0.0	33.3	50.0	11.1	18
Missing							6
Totals	2.3	5.6	10.2	23.6	26.9	31.5	222
6. Informing the public							
Teachers	11.3	13.0	13.6	19.8	14.7	27.7	177
Principals	0.0	0.0	25.0	41.7	33.3	0.0	12
Administrators	0.0	0.0	12.5	50.0	25.0	12.5	8
Others	5.6	11.1	5.6	33.3	22.2	22.2	18
Missing							7
Total	9.8	11.6	13.5	23.3	16.7	25.1	222
7. Policy formation							
Teachers	28.2	20.3	14.1	14.7	10.7	11.9	177
Principals	33.3	8.3	25.0	16.7	8.3	8.3	12
Administrators	0.0	0.0	25.0	37.5	25.0	12.5	8
Others	33.3	11.1	5.6	38.9	5.6	5.6	18
Missing							7
Total	27.9	18.1	14.4	17.7	10.7	11.2	222
8. Reporting to district							
Teachers	20.5	14.8	8.0	12.5	14.8	29.5	176
Principals	8.3	0.0	16.7	33.3	33.3	8.3	12
Administrators	0.0	0.0	25.0	50.0	0.0	25.0	8
Others	16.7	0.0	5.6	44.4	22.2	11.1	18
Missing							8
Total	18.7	12.1	8.9	17.8	15.9	26.6	214

Table 9 (continued)

Percentage ^a of Responses							
Item	0 ^b	1	2	3	4	5	N
9. Student promotion/retention							
Teachers	5.6	34.1	19.6	22.3	8.4	10.1	179
Principals	0.0	25.0	25.0	8.3	33.3	8.3	12
Administrators	12.5	0.0	37.5	12.5	37.5	0.0	8
Others	11.1	22.2	16.7	44.4	5.6	0.0	18
Missing							5
Total	6.0	31.3	20.3	23.0	10.6	8.8	222
10. Teacher evaluation							
Teachers	15.9	27.8	12.5	18.2	11.4	14.2	176
Principals	0.0	25.0	33.3	33.3	8.3	0.0	12
Administrators	25.0	25.0	37.5	0.0	12.5	0.0	8
Others	16.7	16.7	5.6	44.4	11.1	5.6	18
Missing							8
Total	15.4	26.6	14.0	20.6	11.2	12.1	222

^aPercentages do not equal 100% due to rounding

^bResponses were from 1= Never to 3 = Sometimes to 5 = Frequently; 0 = Don't Know

Table 10

Use of Test Scores by Age

	Percentage ^a of Responses						
Item	0 ^b	1	2	3	4	5	N
1. Allocation of funds							
20 - 25	75.0	0.0	0.0	25.0	0.0	0.0	4
26 - 30	46.4	25.0	7.1	14.3	3.6	3.6	28
31 - 35	35.7	28.6	7.1	21.4	7.1	0.0	14
36 - 40	29.2	12.5	20.8	12.5	4.2	20.8	24
41 - 45	32.7	43.6	9.1	3.6	7.3	3.6	55
46 - 50	38.9	33.3	11.1	8.3	5.6	2.8	36
> 50	30.8	51.9	7.7	3.8	3.8	1.9	52
Missing							9
Total	35.7	36.2	9.9	8.5	5.2	4.7	213
2. Assigning students to classes/programs							
20 - 25	0.0	0.0	25.0	25.0	25.0	25.0	4
26 - 30	17.2	20.7	13.8	24.1	10.3	13.8	29
31 - 35	7.1	21.4	7.1	21.4	7.1	35.7	14
36 - 40	7.7	7.7	11.5	23.1	15.4	34.6	26
41 - 45	1.8	9.1	10.9	25.5	16.4	36.4	55
46 - 50	10.8	10.8	5.4	32.4	10.8	29.7	37
> 50	3.7	9.3	3.7	33.3	13.0	37.0	54
Missing							2
Total	6.8	11.4	8.7	27.9	13.2	32.0	219
3. Communicating to parents							
20 - 25	0.0	0.0	25.0	25.0	50.0	0.0	4
26 - 30	3.4	24.1	6.9	34.5	10.3	20.7	29
31 - 35	0.0	21.4	7.1	35.7	28.6	7.1	14
36 - 40	0.0	4.0	16.0	28.0	28.0	24.0	25
41 - 45	3.6	7.3	5.5	34.5	29.1	20.0	55
46 - 50	2.7	10.8	2.7	27.0	29.7	27.0	37
> 50	0.0	5.6	7.4	33.3	18.5	35.2	54
Missing							4
Total	1.8	10.1	7.3	32.1	24.3	24.3	218

Table 10 (continued)

	Percentage ^a of Responses						
Item	0 ^b	1	2	3	4	5	N
4. Curriculum planning							
20 - 25	7.7	0.0	0.0	25.0	50.0	0.0	4
26 - 30	13.8	10.3	13.8	24.1	13.8	24.1	29
31 - 35	7.1	21.4	0.0	14.3	28.6	28.6	14
36 - 40	3.8	3.8	7.7	42.3	23.1	19.2	26
41 - 45	1.9	7.4	3.7	40.7	25.9	20.4	54
46 - 50	2.7	10.8	0.0	32.4	37.8	16.2	37
> 50	7.4	9.3	16.7	22.2	18.5	25.9	54
Missing							4
Total	6.0	9.2	7.8	30.7	24.8	21.6	218
5. Diagnosing skill problems							
20 - 25	25.0	0.0	0.0	25.0	50.0	0.0	4
26 - 30	6.9	3.4	10.3	20.7	31.0	27.6	29
31 - 35	0.0	21.4	0.0	14.3	42.9	21.4	14
36 - 40	4.0	0.0	4.0	28.0	32.0	32.0	25
41 - 45	0.0	3.6	14.5	25.5	23.6	32.7	55
46 - 50	0.0	5.4	8.1	21.6	35.1	29.7	37
> 50	1.9	7.4	13.0	24.1	14.8	38.9	54
Missing							4
Total	2.3	5.5	10.1	23.4	27.1	31.7	218
6. Informing the public							
20 - 25	0.0	0.0	25.0	50.0	25.0	0.0	4
26 - 30	13.8	10.3	6.9	34.5	17.2	17.2	29
31 - 35	0.0	28.6	7.1	21.4	21.4	21.4	14
36 - 40	3.8	15.4	26.9	11.5	11.5	30.8	26
41 - 45	10.9	5.5	12.7	23.6	14.5	32.7	55
46 - 50	13.5	13.5	5.4	21.6	16.2	29.7	37
> 50	9.6	11.5	17.3	23.1	19.2	19.2	52
Missing							5
Total	9.7	11.5	13.4	23.5	16.6	25.3	217

Table 10 (continued)

	Percentage ^a of Responses						
Item	0 ^b	1	2	3	4	5	N
7. Policy formation							
20 - 25	50.0	0.0	0.0	25.0	25.0	0.0	4
26 - 30	31.0	10.3	13.8	27.6	3.4	13.8	29
31 - 35	14.3	21.4	7.1	35.7	7.1	14.3	14
36 - 40	15.4	19.2	15.4	19.2	15.4	15.4	26
41 - 45	34.5	14.5	14.5	14.5	10.9	10.9	55
46 - 50	29.7	24.3	8.1	13.5	13.5	10.8	37
> 50	26.9	21.2	21.2	11.5	9.6	9.6	52
Missing							5
Total	28.1	18.0	14.3	17.5	10.6	11.5	217
8. Reporting to district							
20 - 25	0.0	0.0	0.0	0.0	75.0	25.0	4
26 - 30	31.0	20.7	3.4	6.9	13.8	24.1	29
31 - 35	7.1	14.3	7.1	28.6	28.6	14.3	14
36 - 40	15.4	7.7	3.8	11.5	19.2	42.3	26
41 - 45	21.8	5.5	9.1	18.2	18.2	27.3	55
46 - 50	16.7	13.9	11.1	22.2	11.1	25.0	36
> 50	17.3	15.4	13.5	21.2	7.7	25.0	52
Missing							6
TOTAL	19.0	12.0	8.8	17.6	15.7	26.9	216
9. Student promotion/retention							
20 - 25	0.0	25.0	50.0	0.0	25.0	0.0	4
26 - 30	6.9	41.4	10.3	24.1	13.8	3.4	29
31 - 35	0.0	14.3	28.6	42.9	0.0	14.3	14
36 - 40	3.8	34.6	11.5	30.8	3.8	15.4	26
41 - 45	7.3	36.4	18.2	29.1	3.6	5.5	55
46 - 50	8.1	21.6	21.6	18.9	21.6	8.1	37
> 50	5.6	29.6	25.9	13.0	13.0	13.0	54
Missing							3
Total	5.9	31.1	20.1	25.3	10.5	9.1	219

Table 10 (continued)

Item	Percentage ^a of Responses						N
	0 ^b	1	2	3	4	5	
10. Teacher evaluation							
20 - 25	0.0	0.0	50.0	50.0	0.0	0.0	4
26 - 30	13.8	20.7	24.1	17.2	13.8	10.3	29
31 - 35	7.1	35.7	7.1	21.4	14.3	14.3	14
36 - 40	16.0	8.0	28.0	12.0	8.0	28.0	25
41 - 45	20.0	27.3	7.3	20.0	16.4	9.1	55
46 - 50	10.8	32.4	16.2	21.6	8.1	10.8	37
> 50	19.2	34.6	5.8	23.1	7.7	9.6	52
Missing							6
Total	15.7	26.9	13.9	20.4	11.1	12.0	216

^aPercentages may not equal 100% due to rounding

^bResponses were from 1= Never to 3 = Sometimes to 5 = Frequently; 0 = Don't Know

scores, while the remainder of the respondents indicated that this was a sometimes to frequent use of test scores. As age increased, this became a much more frequent use as indicated by the majority in every group except the 46 to 50 year olds. For informing the public, both the younger group and the older group were again in agreement indicating that this was a sometimes to frequent use of test scores with a larger inclination toward indicating the higher ranking among the 31 to 50 age groups. Of those over 50 years old, only 19.2% said this was a frequent use; the remainder of this group indicated a less frequent use with the majority indicating a sometimes use.

When asked about the use of standardized test scores for policy formation, the older group, greater than 41 years old, indicated that they did not know or that scores were never or sometimes used for this purpose. Of the younger group, the indication was more evenly dispersed, with more of the respondents indicating that test scores were more likely to be used for this purpose. In the area of using test scores to report to the district, the majority of all age groups except the 31 to 35 age group indicated that test scores were frequently used for this purpose. Among all groups, the indications were spread across the scale, indicating perhaps some indecision among the population on this area of the use of standardized test scores. Regarding the use of student test scores for student promotion or retention, the younger age groups, younger than 41, indicated that test scores were sometimes used for that purpose. The three older groups tended more toward never to sometimes in their indications revealing perhaps a more global view of what goes into the promotion or retention of students.

On the important question of using student test scores for teacher evaluation, all age groups except the 36 to 40 age group seemed sure about using test scores for this purpose with a majority in each group indicating never to less than sometimes. Only the 36 to 40 age group indicated sometimes to frequently with 48% giving that ranking. In the oldest group, the majority, 34.6%, stated that student test scores should never be used for teacher evaluation.

When the years of experience of the respondents was matched with standardized test score use, somewhat the same pattern emerges again. The results of this comparison are shown in Table 11. In the area of fund allocation, for example, a majority, 72.1%, of the respondents indicated either that test score data were never used for that purpose, or they indicated that they did not know if test scores were used for that purpose. In the area of assigning students to classes or programs, those with more years' experience were more definite in their indications. While all groups indicated that test scores were sometimes to frequently used for this purpose, the group with greater than 16 years' experience indicated that test scores were more frequently used for assigning students to classes or programs. Those with less experience than that were not as strong in their indications, with more scores in the never to sometimes range. When asked about using test scores to communicate to parents, all groups indicated that this was a sometimes to frequent use of test scores. The group with the least experience, less than five years, were split in that half indicated never to less sometimes, and half indicated sometimes to frequently.

In the area of curriculum planning, there was agreement between the group with fewer years' experience and the more experienced group. Both groups, those with fewer than 16 years' experience and the group with more years' experience stated that test scores were more frequently used for planning purposes than not. When asked about using test scores for diagnosing skill problems among students, all but the group with fewer than 10 years' experience indicated that this was a frequent use of test scores. That group indicated that using students' standardized test scores for diagnosing skill problems was a sometimes to less than frequent utilization. As stated earlier, this use was the most frequently-named use of student test scores among all respondents. When contrasted with years' experience, however, it was also the most frequent use among those with greater than 11 years of experience.

For the use of test scores to inform the public, a majority in the middle of the

Table 11

Use of Test Scores by Years' Experience

	Percentage ^a of Responses						
Item	0 ^b	1	2	3	4	5	N
1. Allocation of funds							
< 5	57.9	26.3	0.0	10.5	0.0	5.3	19
5 - 10	48.6	20.0	5.7	14.3	5.7	5.7	35
11 - 15	35.3	26.5	17.6	5.9	5.9	8.8	34
16 - 20	33.3	39.4	6.1	12.1	6.1	3.0	33
21 - 25	29.3	41.5	12.2	4.9	7.3	4.9	41
26 - 30	25.0	53.6	10.7	7.1	0.0	3.6	28
> 30	27.3	50.0	9.1	4.5	9.1	0.0	22
Missing							10
Total	35.8	36.3	9.4	8.5	5.2	4.7	222
2. Assigning students to classes/programs							
< 5	10.0	25.0	10.0	20.0	15.0	20.0	20
5 - 10	16.2	16.2	10.8	21.6	8.1	27.0	37
11 - 15	2.9	11.8	14.7	35.3	2.9	32.4	34
16 - 20	3.0	6.1	9.1	30.3	24.2	27.3	33
21 - 25	4.8	9.5	2.4	26.2	21.4	35.7	42
26 - 30	7.1	0.0	3.6	42.9	3.6	42.9	28
> 30	4.3	17.4	8.7	17.4	13.0	39.1	23
Missing							5
Total	6.9	11.5	8.3	28.1	12.9	32.3	222
3. Communicating to parents							
< 5	0.0	35.0	15.0	20.0	20.0	10.0	20
5 - 10	2.8	11.1	2.8	47.2	13.9	22.2	36
11 - 15	0.0	8.8	8.8	35.3	26.5	20.6	34
16 - 20	0.0	3.0	6.1	30.3	33.3	27.3	33
21 - 25	7.1	9.5	4.8	26.2	28.6	23.8	42
26 - 30	0.0	3.6	3.6	39.3	17.9	35.7	28
> 30	0.0	8.7	13.0	21.7	30.4	26.1	23
Missing							6
Total	1.9	10.2	6.9	32.4	24.5	24.1	222

Table 11 (continued)

	Percentage ^a of Responses						
Item	0 ^b	1	2	3	4	5	N
4. Curriculum planning							
< 5	10.5	10.5	10.5	15.8	36.8	15.8	19
5 - 10	16.2	10.8	8.1	43.2	2.7	18.9	37
11 - 15	0.0	8.8	2.9	32.4	26.5	29.4	34
16 - 20	3.0	6.1	3.0	30.3	33.3	24.2	33
21 - 25	2.4	11.9	7.1	28.6	31.0	19.0	42
26 - 30	3.6	10.7	7.1	28.6	28.6	21.4	28
> 30	8.7	4.3	21.7	21.7	21.7	21.7	23
Missing							6
Total	6.0	9.3	7.9	30.1	25.0	21.8	222
5. Diagnosing skill problems							
< 5	5.0	5.0	10.0	10.0	45.0	25.0	20
5 - 10	5.4	5.4	10.8	40.5	21.6	16.2	37
11 - 15	0.0	5.9	11.8	17.6	23.5	41.2	34
16 - 20	3.0	3.0	6.1	24.2	27.3	36.4	33
21 - 25	0.0	7.1	9.5	19.0	33.3	31.0	42
26 - 30	0.0	10.7	7.1	25.0	17.9	39.3	28
> 30	4.3	0.0	17.4	21.7	21.7	34.8	23
Missing							5
Total	2.3	5.5	10.1	23.5	26.7	31.8	222
6. Informing the public							
< 5	15.0	15.0	10.0	35.0	10.0	15.0	20
5 - 10	8.1	21.6	8.1	27.0	16.2	18.9	37
11 - 15	5.9	8.8	8.8	17.6	20.6	38.2	34
16 - 20	18.2	6.1	24.2	18.2	6.1	27.3	33
21 - 25	11.9	9.5	16.7	19.0	21.4	21.4	42
26 - 30	0.0	14.8	7.4	25.9	14.8	37.0	27
> 30	9.1	4.5	18.2	31.8	22.7	13.6	22
Missing							7
Total	9.8	11.6	13.5	23.7	16.3	25.1	222

Table 11 (continued)

	Percentage ^a of Responses						
Item	0 ^b	1	2	3	4	5	N
7. Policy formation							
< 5	35.0	15.0	0.0	30.0	5.0	15.0	20
5 - 10	43.2	16.2	10.8	16.2	5.4	8.1	37
11 - 15	26.5	14.7	17.6	20.6	8.8	11.8	34
16 - 20	24.2	12.1	6.1	24.2	12.1	21.2	33
21 - 25	16.7	23.8	21.4	16.7	16.7	4.8	42
26 - 30	23.1	26.9	15.4	15.4	7.7	11.5	26
> 30	30.4	17.4	26.1	0.0	17.4	8.7	23
Missing							7
Total	27.9	18.1	14.4	17.7	10.7	11.2	222
8. Reporting to district							
< 5	20.0	25.0	0.0	10.0	20.0	25.0	20
5 - 10	32.4	16.2	2.7	18.9	13.5	16.2	37
11 - 15	14.7	5.9	8.8	14.7	20.6	35.3	34
16 - 20	15.2	0.0	9.1	24.2	21.2	30.3	33
21 - 25	19.0	11.9	14.3	16.7	16.7	21.4	42
26 - 30	7.7	15.4	15.4	23.1	3.8	34.6	26
> 30	18.2	18.2	9.1	13.6	13.6	27.3	22
Missing							8
Total	18.7	12.1	8.9	17.8	15.9	26.6	222
9. Student promotion/retention							
< 5	0.0	45.0	20.0	10.0	15.0	10.0	20
5 - 10	8.1	32.4	16.2	35.1	5.4	2.7	37
11 - 15	2.9	32.4	17.6	23.5	11.8	11.8	34
16 - 20	12.1	30.3	18.2	21.2	9.1	9.1	33
21 - 25	7.1	28.6	23.8	28.6	7.1	4.8	42
26 - 30	0.0	21.4	28.6	17.9	14.3	17.9	28
> 30	8.7	30.4	17.4	13.0	17.4	13.0	23
Missing							5
Total	6.0	30.9	20.3	23.0	10.6	9.2	222

Table 11 (continued)

Item	Percentage ^a of Responses						N
	0 ^b	1	2	3	4	5	
10. Teacher evaluation							
< 5	0.0	35.0	25.0	20.0	15.0	5.0	20
5 - 10	16.7	22.2	13.9	25.0	2.8	19.4	36
11 - 15	17.6	14.7	14.7	20.6	20.6	11.8	34
16 - 20	15.6	28.1	6.3	21.9	6.3	21.9	32
21 - 25	26.2	31.0	14.3	11.9	11.9	4.8	42
26 - 30	7.1	32.1	14.3	25.0	7.1	14.3	28
> 30	18.2	31.8	9.1	22.7	13.6	4.5	22
Missing							8
Total	15.9	27.1	13.6	20.6	10.7	12.1	222

^aPercentages may not equal 100% due to rounding

^bResponses were from 1= Never to 3 = Sometimes to 5 = Frequently; 0 = Don't Know

range from greater than 10 years' experience to fewer than 30 years' experience indicated that this was a more frequent use for test score data. Those with fewer than five years' experience and those with greater than 30 years felt that this was a sometimes use of test score data. Those in the group from 21 to 25 years' experience, almost an equal number, were split in their ranks, with some indicating this purpose was either a frequent use or only a sometimes use. Once again, those with less experience viewed the use of test scores differently than those with more experience.

The scores for the use of test data for policy formation also showed a split among three of the groups. Those with fewer years of experience, fewer than 16, indicated they were unsure about this use by the majority within each group indicating they did not know if test scores were used for this purpose. Those in the middle group, from 16 to 20 years' experience, indicated either that this was a sometimes use or that they did not know. Those with greater than 25 years' experience also indicated uncertainty or indicated that test scores were never used for this purpose. By years' experience, more respondents indicated uncertainty, by responding that they did not know, on this item than any others.

The use of test score data for reporting to the district was indicated as a frequent use by the majority of the respondents in all groups except those with fewer than 11 years' experience. A dichotomy existed in the group with fewer than five years' experience, as 25% responded never, and 25% responded frequently. Among those with 5 to 10 years' experience, the majority, 32.4%, indicated they did not know. Those with greater than 30 years' experience, however, were not as definitive as the others, spreading their marks across the entire range from don't know to frequently. In the area of student promotion or retention, the majority of the respondents indicated that this was a very infrequent use, with many respondents in each group indicating that student test scores were never used for this purpose. The group with 26 to 30 years' experience was less sure than the others, however, with scores that were fairly evenly

divided across the range of scores.

When the question of using student test data for teacher evaluation was posed, the responses compared with years' experience indicated much the same results as they did when compared with the respondents' age. A majority of those with fewer than five years' experience indicated that student test scores should never be used for the purpose of teacher evaluation. A majority of those with between 5 and 15 years' indicated that could be a more frequent use than those with fewer years' experience. A majority of those with more years' experience, greater than 16, indicated that this should never be a use for student test data. Not surprisingly, those respondents who were older with more years' experience indicated they were against the use of student test score data for teacher evaluation. Those respondents, like others of their age and experience level, appeared to be more resistant to change than their younger, less-experienced colleagues.

The next series of items asked the respondents how frequently they discussed standardized test results with various entities, from the District Office to individual students. This survey item also dealt with the use of standardized test results in that discussions with others indicated the degree to which an individual uses the terms associated with testing and the results of those tests as a means of communicating. The higher the degree of communication, the more use the results find among educators. The results are shown in Table 12. The most common discussion of standardized test scores takes place with other classroom teachers; the least discussion takes place between educators and the district office. As Dorr-Bremme and Herman (1986) suggested, teachers communicate more with those they are closest to in their workplace. Consequently, teachers would tend to discuss test scores with other teachers more, especially when the discussion of students or areas of difficulty arises. Because teachers do not visit the district office frequently, it not being their area of influence, they would only discuss test scores with those members of the staff that directly impact them: i.e., a special education teacher would discuss scores with the special education

Table 12

Discussion of Test Score Results with Others

Item	Responses	
	N	Mean ^a
Other Classroom Teachers	221	3.48
Principal	218	3.19
Parents	221	3.18
Guidance Department	218	2.57
Students	220	2.51
District Office	219	1.84

^aResponses were from 1= Never to 5 = Frequently with 3 = Sometimes

supervisor, or a secondary science teacher might discuss ACT or SAT science test scores with the district's science director.

It is interesting that teachers do not discuss test scores to a great degree with their students or their school's guidance department, those scores ranking just above discussions with district office personnel. If tests are to help the student ascertain his level of knowledge or assess strengths and weaknesses, then it would seem that teachers would be interested in communicating that information to their students. Perhaps while they do not communicate test information, they do communicate information about the student's progress or problem areas in other ways. According to the results of this survey, it seems that teachers do not discuss test scores with students; rather, they communicate more to the principal and the parents. By the same token, it appears that teachers do not communicate test score information to the school's guidance counselors. It may be that guidance counselors perform the communication in reverse, giving the information to the teachers rather than the other way around.

The item dealing with various bases for educational decisions and the degree respondents preferred to use those bases bears direct relation to the use of test scores. As educators make their decisions about students and programs, they utilize varied sources of information. Standardized test scores are just one of those sources. The results from this item are shown in Table 13. Not surprisingly, most teachers felt that their own observations and students' classwork were the best source of information for making decisions regarding their students. As Herman and Dorr-Bremme (1984) suggested, teachers feel their own assessment procedures are better than non-teacher designed methods. Because teachers observe the students many times over a long period of time, use a variety of assessment methods, and call into play some sense called "intuition," they sense they can perceive more about the student than an impersonal, standardized test given only at one time on one day where the questions have only one correct answer (Bracey, 1992; Darling-Hammond, 1991; Shepard, 1989).

Table 13

Preferences for Bases of Educational Decisions

Items	Responses	
	N	Mean ^a
Own observations and students' classwork	221	4.61
Results of teacher-made tests	221	3.92
Results of other special placement tests	221	3.02
Curriculum materials' placement test results	220	3.01
Previous teachers' comments, reports, grades	221	2.97
Students' scores of T-CAP tests	220	2.74
Students' other standardized test scores	220	2.57

^aResponses were from 1= Never to 5 = Frequently with 3 = Sometimes

Concurrently, the respondents also gave high marks to the use of the teacher-made test as a component in the decision-making process. Following the same reasoning as before, it appeared that teachers tended to place more emphasis on what they had devised as an assessment instrument rather than on an instrument designed by one who had never seen the student or who had never been in the classroom with the teacher and the student. The responses that received the lowest scores were those that referred explicitly to standardized tests and the state-mandated T-CAP test. This indicated that practitioners had little regard for standardized tests as a major component in their decision-making process. The use of curriculum materials or other non-teacher-made materials and results of other special placement tests appeared to fall in the middle of this spectrum. This indicated some reliance on course materials, but not as much as on teachers' materials.

When respondents' years of experience were compared to their educational decisions, shown in Table 14, one strong response emerged in the area of using their own observations and students' work as a basis for decision-making, reinforcing the above findings. For the first statement, basing decisions on curriculum materials' placement test results, the majority of all groups except those with fewer than five years' experience stated they sometimes used this as a basis for their decisions. The group with the least experience showed less reliance on this information, with half the number responding indicating more than sometimes. Using curriculum materials' placement test results was not a strong basis for educational decisions among the respondents.

The next statement in this item, using one's own observation and students' classwork as a basis for educational decisions, found the greatest agreement among all groups. Based on previous findings, this was not unexpected. A majority in each group stated that this was the preferred basis for making their educational decisions by indicating they frequently used their own observations and students' classwork. Only a few in each group indicated less than a frequent use of this information. For the next

Table 14

Bases for Educational Decisions by Years' Experience

	Percentage ^a of Responses					
Items	1 ^b	2	3	4	5	N
1. Curriculum materials' placement test results:						
< 5	10.0	30.0	25.0	25.0	10.0	20
5 - 10	13.5	16.2	51.4	16.2	2.7	37
11 - 15	8.6	22.9	48.6	11.4	8.6	35
16 - 20	9.1	18.2	54.5	12.1	6.1	33
21 - 25	11.6	11.6	46.5	18.6	11.6	43
26 - 30	14.8	14.8	29.6	22.2	18.5	27
> 30	0.0	17.4	39.1	26.1	17.4	23
Missing						4
Total	10.1	17.9	44.0	17.9	10.1	222
2. Own observations and students' classwork:						
< 5	0.0	0.0	0.0	30.0	70.0	20
5 - 10	0.0	0.0	2.7	16.2	81.1	37
11 - 15	0.0	2.9	5.7	14.3	77.1	35
16 - 20	0.0	0.0	9.1	18.2	72.7	33
21 - 25	4.7	0.0	4.7	14.0	76.7	43
26 - 30	3.6	7.1	10.7	14.3	64.3	28
> 30	0.0	4.3	4.3	17.4	73.9	23
Missing						3
Total	1.4	1.8	5.5	16.9	74.4	222
3. Previous teachers' comments, reports, grades:						
< 5	10.0	35.0	40.0	15.0	0.0	20
5 - 10	13.5	27.0	32.4	21.6	5.4	37
11 - 15	11.4	17.1	42.9	20.0	8.6	35
16 - 20	12.1	30.3	33.3	18.2	6.1	33
21 - 25	7.0	20.9	37.2	23.3	11.6	43
26 - 30	7.1	17.9	39.3	10.7	25.0	28
> 30	8.7	17.4	21.7	26.1	26.1	23
Missing						3
Total	10.0	23.3	35.7	19.6	11.4	222

Table 14 (continued)

	Percentage ^a of Responses					
Items	1 ^b	2	3	4	5	N
4. Results of other special placement tests:						
< 5	2.0	50.0	25.0	20.0	0.0	20
5 - 10	13.5	13.5	45.9	24.3	2.7	37
11 - 15	2.9	22.9	51.4	17.1	5.7	35
16 - 20	12.1	21.2	39.4	18.2	9.1	33
21 - 25	7.0	11.6	41.9	32.6	7.0	43
26 - 30	14.3	7.1	35.7	25.0	17.9	28
> 30	8.7	13.0	30.4	39.1	8.7	23
Missing						3
Total	9.1	18.3	40.2	25.1	7.3	222
5. Results of teacher-made tests:						
< 5	0.0	10.0	20.0	35.0	35.0	20
5 - 10	5.4	10.8	27.0	32.4	24.3	37
11 - 15	0.0	2.9	20.0	37.1	40.0	35
16 - 20	6.1	9.1	18.2	30.3	36.4	33
21 - 25	4.7	2.3	16.3	34.9	41.9	43
26 - 30	3.6	7.1	32.1	17.9	39.3	28
> 30	0.0	8.7	17.4	43.5	30.4	23
Missing						3
Total	3.2	6.8	21.5	32.9	35.6	222
6. Students' other standardized test results:						
< 5	21.1	52.6	15.8	10.5	0.0	19
5 - 10	35.1	29.7	24.3	10.8	0.0	37
11 - 15	14.3	31.4	37.1	14.3	2.9	35
16 - 20	27.3	24.2	33.3	6.1	9.1	33
21 - 25	14.0	27.9	30.2	14.0	14.0	43
26 - 30	10.7	25.0	39.3	17.9	7.1	28
> 30	13.0	21.7	34.8	21.7	8.7	23
Missing						4
Total	19.7	29.4	31.2	13.3	6.4	222

Table 14 (continued)

	Percentage ^a of Responses					
Items	1 ^b	2	3	4	5	N
7. Students' scores of T-CAP tests:						
< 5	10.5	57.9	15.8	5.3	10.5	19
5 - 10	32.4	24.3	29.7	13.5	0.0	37
11 - 15	20.0	28.6	31.4	11.4	8.6	35
16 - 20	21.2	12.1	45.5	9.1	12.1	33
21 - 25	9.3	27.9	32.6	25.6	4.7	43
26 - 30	14.3	14.3	42.9	14.3	14.3	28
> 30	4.3	30.4	13.0	30.4	21.7	23
Missing						4
Total	17.0	26.1	31.7	16.1	9.2	222

^aPercentages do not equal 100% due to rounding

^bResponses were from 1= Never to 5 = Frequently, with 3 = Sometimes

statement, using previous teachers' comments, reports, or grades as a basis for decisions, all groups except the group with the most experience indicated that this information was used only sometimes as a basis for their decisions. A majority of the respondents in the group with the most experienced indicated that this information was used frequently or more than sometimes, a departure from the rest of the respondents. While this was not a clear-cut majority, it was interesting to see the most-experienced group base their decisions on such subjective information.

The next statement dealt with using results of other special placement tests as a basis for decision making. Again, this information was indicated as a basis sometimes by all groups, with the a majority in the most-experienced group indicating slightly more than sometimes. A majority in the least-experienced group indicated a ranking slightly less than sometimes, but overall, it, too, indicated that it used this information only sparingly to make decisions. The statement dealing with using the results of teacher-made tests as a basis for decision-making again saw a strong frequently response from among all groups, with the weakest response from those with fewer than 10 years' experience and the group with greater than 30 years'. The groups indicated that this was a frequently-used basis for their educational decisions, much like the earlier statement concerning their own observations and students' classwork.

The next two statements in this item asked about the use of standardized tests and T-CAP tests specifically as a basis for educational decisions. These statements were posed last in the list as a means of isolating them from the other items. The previous statements did not deal with standardized tests of any sort, while these were aimed specifically at the respondents' preferences for using standardized test data as a basis for educational decisions. For both statements, the majority of all groups indicated that these scores were only used sometimes or less as a basis for making educational decisions. A slight difference existed among those with greater than 30 years' experience between using standardized test scores and T-CAP scores. This group

indicated using the T-CAP slightly more as a basis for making decisions than did the other groups. A majority in the group with 5 to 10 years' experience indicated that they never used either a standardized test or the T-CAP as much as the other groups by a narrow margin. Only two people in the least-experienced group used the T-CAP frequently as a basis for making decisions. The relatively low scores of the use of both standardized tests and the T-CAP test showed how educators in Tennessee felt about these tests as bases for educational decisions.

Perceptions of Testing Terms and Term Usage

A major part of the study was involved in what information educators possessed about standardized tests, testing procedures, and how--or if--that information was used. One way to gain this information would be to ask questions about standardized tests in a multiple-choice format as did VanLeirsburg and Johns in their 1991 study. Another way would be to interview each of the respondents to ascertain their level of knowledge and term usage as did Brown in his 1991 study. Still another way would be to inquire of the respondents how well they could define a term associated with standardized testing and how often they used that term with others. Because the intent of this study was to determine the *perceptions* of levels of knowledge of educators rather than attempt to determine exactly their level of knowledge, the latter way was selected. Respondents were asked how well they could explain a term to someone who was unfamiliar with testing and measurement. The terms selected were those normally associated with tests and measurement. In addition, respondents were asked how frequently they used those terms in conversations with others, i.e. other teachers, administrators, parents, or students. The responses for the definition section were measured on a five point scale from "Not Well" to "Very Well" with a mid-range value for "Fairly Well." A response of "Don't Know" was also provided. The responses for the level of use portion were

measured on a five point scale from "Never Use" to "Frequently Use." There was no provision for a "Don't Know." The number of respondents and the means of the responses for both questions are shown in Table 15. The responses are listed in rank order of the means of the perceptions of the term's meaning.

The number of responses for each term varies. In 14 of 17 instances, the same number did not respond to the question about term use that responded to term definition. This suggests that some respondents indicated that while they could define a term in some respects, they indicated they did not use it in conversations with others. While this may display a tendency for educators to say they know a term but are unsure about it, it probably represents an attempt to provide a "Don't Know" response when one was not offered.

The most-understood term was percentile with percentile rank the next most-understood term. As stated, no attempt was made to determine if the respondents actually could *define* the term, only if they *perceived* they understood what it meant. According to these responses, the least-understood term was statistical process control, a term unfamiliar to many educators. The next least-understood term was adjusted score, which was unusual because an adjusted score is the basis for determining the percentile and percentile rank for the placing of individual scores within a range of scores. If those surveyed truly understood percentile and percentile rank, they should also comprehend the concept of the adjusted score. Perhaps they did understand the concept, but not the term. The gap between perception of what is thought to be known and what really is known by educators on this subject may be wider than originally thought.

The term scale score, an important concept in the science of tests and measurement, appeared near the bottom of the rankings, displaying a lack of perception of a significant expression. As scores are converted from raw scores using a variety of transformations, they eventually become scaled scores, the scores that are used for all other manipulations. A low understanding of this term reinforced the low perception

Table 15

Perceptions and Use of Testing Terms

Terms	Responses				
	Term Perception		Term Use		Difference a-b
	N ^a	Mean ^b	N ^a	Mean ^c	
Percentile	212	4.37	210	3.20	1.17
Percentile rank	213	4.34	211	3.03	1.31
Raw score	213	4.16	209	2.42	1.74
Mean	211	4.13	211	2.47	1.66
Median	213	4.08	209	2.31	1.77
Grade equivalents (GE)	212	4.00	211	2.69	1.31
Stanine	212	3.93	209	2.93	1.00
Norm-referenced test	211	3.86	209	2.21	1.65
Mode	212	3.82	211	2.11	1.71
Normal distribution	212	3.70	210	2.06	1.64
Standard Score	210	3.58	209	2.33	1.25
Reliability	213	3.51	210	2.28	1.23
Criterion-referenced test	209	3.48	209	2.16	1.32

Table 15 (continued)

Terms	Responses				
	Term Perception		Term Use		Difference a-b
	N ^a	Mean ^b	N ^a	Mean ^c	
Validity	213	3.35	211	2.18	1.17
Scale Score	212	3.33	208	2.06	1.27
Adjusted Score	211	2.70	211	1.62	1.08
Statistical process control	212	2.38	211	1.44	0.94

^aNumber responding (of the 222 participants)

^bResponses were from 1= Not Well to 5 = Very Well with 3 = Fairly Well

^cResponses were from 1= Never Use to 5 = Sometimes Use with 3 = Frequently Use

level as previously stated. The few number of "Don't Know" responses also pointed out the educators' perceptions that they "ought" to know what a term means whether they did or not.

The difference between the knowing of a term and its use was significant. In all cases, the stated ability to define a term had a greater mean than the term's use, in some cases, quite a large margin. This suggested that while the respondents may think they know what a term means, they do not use it, perhaps because the opportunity to engage in conversations that utilize the terms occurs very infrequently. The two terms that respondents said were not used frequently were adjusted score and statistical process control. It is rare that the latter term would come up in conversations about standardized tests, but the other could be used more frequently when discussing students and tests.

The perception of the terms listed in this item were compared with the respondents' age and years of experience. It was thought that those educators who were older and had more years' experience would be more familiar with testing and its terms due to that age and experience level and would thus have more knowledge of testing terms. The results of those comparisons are shown in Table 16 and Table 17.

As age increased, so did the incidence of knowledge of testing terms. Figure 1 shows a comparison of the percentages of those respondents who indicated they could explain a term very well or slightly less (a ranking of 4 or 5) between the groups of from 20 to 35 years of age and those greater than 36 years old. In all cases the stated ability to define the terms associated with testing was greater as one's age increased. Very few respondents indicated they did not know a term, except for the last term, statistical process control, and most indicated they could not explain that term to another with confidence. As the years' experience increased, the results were not so conclusive. With an increase in experience, the ability to explain a term to another was not as pronounced as it was with an increase in age. Figure 2 shows a comparison of the

Table 16

Perceptions of Testing Terms by Age

	Percentage ^a of Responses						
Items	0 ^b	1	2	3	4	5	N
1. Adjusted Score:							
20 - 25	25.0	0.0	0.0	75.0	0.0	0.0	4
26 - 30	7.1	10.7	25.0	32.1	21.4	3.6	28
31 - 35	8.3	25.0	25.0	41.7	0.0	0.0	12
36 - 40	0.0	16.7	29.2	33.3	8.3	12.5	24
41 - 45	0.0	27.3	21.8	29.1	18.2	3.6	55
46 - 50	8.3	13.9	22.2	36.1	5.6	13.9	36
> 50	3.8	17.3	15.4	34.6	17.3	11.5	52
Missing							11
Total	4.3	18.5	21.3	34.1	13.7	8.1	222
2. Criterion-referenced test:							
20 - 25	0.0	0.0	25.0	75.0	0.0	0.0	4
26 - 30	3.6	3.6	17.9	35.7	21.4	17.9	28
31 - 35	0.0	8.3	0.0	16.7	41.7	33.3	12
36 - 40	0.0	4.3	8.7	39.1	21.7	26.1	23
41 - 45	0.0	10.9	1.8	40.0	29.1	18.2	55
46 - 50	5.7	2.9	0.0	34.3	31.4	25.7	35
> 50	3.8	13.5	13.5	23.1	19.2	26.9	52
Missing							13
Total	2.4	8.1	7.7	33.5	25.4	23.0	222
3. Grade equivalents (GE):							
20 - 25	0.0	25.0	0.0	25.0	50.0	0.0	4
26 - 30	10.3	3.4	13.8	13.8	27.6	31.0	29
31 - 35	0.0	0.0	0.0	8.3	8.3	83.3	12
36 - 40	0.0	0.0	0.0	24.0	36.0	40.0	25
41 - 45	1.9	5.6	3.7	14.8	33.3	40.7	54
46 - 50	0.0	0.0	2.8	25.0	33.3	38.9	36
> 50	1.9	5.8	9.6	17.3	23.1	42.3	52
Missing							10
Total	2.4	3.8	5.7	17.9	29.2	41.0	222

Table 16 (continued)

Percentage ^a of Responses							
Items	0 ^b	1	2	3	4	5	N
4. Mean:							
20 - 25	0.0	0.0	0.0	0.0	25.0	75.0	4
26 - 30	0.0	0.0	3.4	24.1	24.1	48.3	29
31 - 35	0.0	0.0	0.0	0.0	25.0	75.0	12
36 - 40	0.0	0.0	0.0	28.0	40.0	32.0	25
41 - 45	0.0	1.8	5.5	23.6	21.8	47.3	55
46 - 50	0.0	0.0	0.0	13.9	38.9	47.2	36
> 50	2.0	4.0	14.0	16.0	20.0	44.0	50
Missing							11
Total	0.5	1.4	5.2	19.0	27.0	46.9	222
5. Median:							
20 - 25	0.0	0.0	0.0	0.0	25.0	75.0	4
26 - 30	0.0	0.0	6.9	24.1	24.1	44.8	29
31 - 35	0.0	0.0	0.0	0.0	25.0	75.0	12
36 - 40	0.0	0.0	8.0	28.0	32.0	32.0	25
41 - 45	0.0	1.8	5.5	21.8	23.6	47.3	55
46 - 50	0.0	0.0	0.0	16.7	33.3	50.0	36
> 50	1.9	5.8	9.6	25.0	13.5	44.2	52
Missing							9
Total	0.5	1.9	5.6	21.1	23.9	46.9	222
6. Mode:							
20 - 25	0.0	0.0	0.0	0.0	25.0	75.0	4
26 - 30	0.0	0.0	10.3	24.1	24.1	41.4	29
31 - 35	0.0	8.3	0.0	16.7	25.0	50.0	12
36 - 40	0.0	8.0	12.0	32.0	20.0	28.0	25
41 - 45	0.0	3.6	9.1	25.5	21.8	40.0	55
46 - 50	0.0	0.0	2.9	25.7	31.4	40.0	35
> 50	1.9	11.5	13.5	15.4	23.1	34.6	52
Missing							10
Total	0.5	5.2	9.0	22.6	24.1	38.7	222

Table 16 (continued)

Percentage ^a of Responses							
Items	0 ^b	1	2	3	4	5	N
7. Normal distribution:							
20 - 25	0.0	0.0	0.0	75.0	0.0	25.0	4
26 - 30	0.0	6.9	3.4	27.6	27.6	34.5	29
31 - 35	0.0	0.0	0.0	16.7	41.7	41.7	12
36 - 40	0.0	4.0	8.0	32.0	40.0	16.0	25
41 - 45	3.6	1.8	12.7	27.3	21.8	32.7	55
46 - 50	0.0	5.7	8.6	28.6	31.4	25.7	35
> 50	0.0	7.7	11.5	26.9	21.2	32.7	52
Missing							10
Total	0.9	4.7	9.0	28.3	26.9	30.2	222
8. Norm-referenced test:							
20 - 25	0.0	0.0	25.0	25.0	50.0	0.0	4
26 - 30	0.0	3.4	3.4	27.6	34.5	31.0	29
31 - 35	0.0	0.0	0.0	0.0	33.3	66.7	12
36 - 40	0.0	0.0	8.0	20.0	48.0	24.0	25
41 - 45	1.8	5.5	3.6	29.1	23.6	36.4	55
46 - 50	0.0	2.9	5.7	25.7	31.4	34.3	35
> 50	2.0	7.8	11.8	17.6	25.5	35.3	51
Missing							11
Total	0.9	4.3	6.6	22.7	30.8	34.6	222
9. Percentile:							
20 - 25	0.0	0.0	0.0	25.0	50.0	25.0	4
26 - 30	0.0	3.4	0.0	17.2	34.5	44.8	29
31 - 35	0.0	0.0	0.0	0.0	25.0	75.0	12
36 - 40	0.0	0.0	0.0	36.0	28.0	36.0	25
41 - 45	0.0	1.8	0.0	10.9	36.4	50.9	55
46 - 50	0.0	0.0	0.0	2.8	44.4	52.8	36
> 50	0.0	2.0	7.8	7.8	25.5	56.9	51
Missing							10
Total	0.0	1.4	1.9	12.3	33.5	50.9	222

Table 16 (continued)

Percentage ^a of Responses							
Items	0 ^b	1	2	3	4	5	N
10. Percentile rank:							
20 - 25	0.0	0.0	0.0	25.0	50.0	25.0	4
26 - 30	0.0	3.4	3.4	17.2	31.0	44.8	29
31 - 35	0.0	0.0	0.0	8.3	16.7	75.0	12
36 - 40	0.0	4.0	0.0	28.0	28.0	40.0	25
41 - 45	0.0	1.8	1.8	12.7	29.1	54.5	55
46 - 50	0.0	0.0	0.0	5.6	41.7	52.8	36
> 50	0.0	1.9	9.6	11.5	23.1	53.8	52
Missing							9
Total	0.0	1.9	3.3	13.6	29.6	51.6	222
11. Raw score:							
20 - 25	0.0	0.0	25.0	50.0	0.0	25.0	4
26 - 30	0.0	3.4	6.9	31.0	20.7	37.9	29
31 - 35	0.0	0.0	0.0	8.3	25.0	66.7	12
36 - 40	0.0	12.0	12.0	32.0	16.0	28.0	25
41 - 45	0.0	3.6	1.8	21.8	29.1	43.6	55
46 - 50	0.0	0.0	2.8	25.0	27.8	44.4	36
> 50	0.0	3.8	5.8	13.5	23.1	53.8	52
Missing							9
Total	0.0	3.8	5.2	22.5	23.9	44.6	222
12. Reliability:							
20 - 25	0.0	0.0	50.0	25.0	25.0	0.0	4
26 - 30	3.4	3.4	10.3	24.1	34.5	24.1	29
31 - 35	0.0	0.0	0.0	33.3	41.7	25.0	12
36 - 40	0.0	12.0	20.0	36.0	24.0	8.0	25
41 - 45	3.6	9.1	9.1	30.9	25.5	21.8	55
46 - 50	0.0	2.8	13.9	25.0	33.3	25.0	36
> 50	1.9	7.7	13.5	21.2	21.2	34.6	52
Missing							9
Total	1.9	6.6	12.7	27.2	27.7	23.9	222

Table 16 (continued)

	Percentage ^a of Responses						
Items	0 ^b	1	2	3	4	5	N
13. Scale Score:							
20 - 25	0.0	0.0	50.0	25.0	25.0	0.0	4
26 - 30	6.9	6.9	20.7	20.7	24.1	20.7	29
31 - 35	0.0	8.3	25.0	16.7	33.3	16.7	12
36 - 40	0.0	12.0	16.0	32.0	24.0	16.0	25
41 - 45	3.6	12.7	5.5	34.5	27.3	16.4	55
46 - 50	0.0	2.9	17.1	37.1	14.3	28.6	35
> 50	5.8	9.6	21.2	13.5	23.1	26.9	52
Missing							10
Total	3.3	9.0	16.5	26.4	23.6	21.2	222
14. Standard Score:							
20 - 25	0.0	0.0	25.0	50.0	25.0	0.0	4
26 - 30	0.0	0.0	13.8	34.5	24.1	27.6	29
31 - 35	0.0	0.0	25.0	16.7	41.7	16.7	12
36 - 40	0.0	13.0	17.4	21.7	21.7	26.1	23
41 - 45	1.9	5.6	7.4	22.2	37.0	25.9	54
46 - 50	0.0	2.8	5.6	47.2	13.9	30.6	36
> 50	3.8	7.7	19.2	11.5	26.9	30.8	52
Missing							12
Total	1.4	5.2	13.3	25.7	27.1	27.1	222
15. Stanine:							
20 - 25	0.0	0.0	25.0	50.0	25.0	0.0	4
26 - 30	3.4	3.4	6.9	20.7	44.8	20.7	29
31 - 35	8.3	16.7	0.0	16.7	16.7	41.7	12
36 - 40	0.0	8.3	4.2	25.0	33.3	29.2	24
41 - 45	1.8	3.6	5.5	14.5	36.4	38.2	55
46 - 50	0.0	0.0	5.6	16.7	30.6	47.2	36
> 50	0.0	7.7	7.7	11.5	26.9	46.2	52
Missing							10
Total	1.4	5.2	6.1	17.0	32.5	37.7	222

Table 16 (continued)

Items	Percentage ^a of Responses						N
	0 ^b	1	2	3	4	5	
16. Statistical process control:							
20 - 25	0.0	75.0	0.0	25.0	0.0	0.0	4
26 - 30	13.8	24.1	20.7	20.7	10.3	10.3	29
31 - 35	0.0	25.0	33.3	16.7	8.3	16.7	12
36 - 40	4.0	44.0	12.0	28.0	8.0	4.0	25
41 - 45	10.9	30.9	18.2	25.5	10.9	3.6	55
46 - 50	11.4	31.4	25.7	17.1	2.9	11.4	35
> 50	9.6	26.9	17.3	25.0	9.6	11.5	52
Missing							10
Total	9.4	31.1	19.3	23.1	8.5	8.5	222
17. Validity:							
20 - 25	0.0	0.0	75.0	0.0	25.0	0.0	4
26 - 30	3.4	6.9	10.3	24.1	24.1	31.0	29
31 - 35	0.0	8.3	25.0	16.7	25.0	25.0	12
36 - 40	0.0	20.0	20.0	24.0	24.0	12.0	25
41 - 45	5.5	10.9	14.5	21.8	23.6	23.6	55
46 - 50	0.0	8.3	13.9	22.2	27.8	27.8	36
> 50	0.0	13.5	13.5	21.2	28.8	23.1	52
Missing							9
Total	1.9	11.3	16.0	21.6	25.8	23.5	222

^aPercentages may not equal 100% due to rounding

^bResponses were from 1= Not Well to 3 = Fairly Well to 5 = Very Well, with 0 = Don't Know

Table 17

Perceptions of Testing Terms by Years' Experience

	Percentage ^a of Responses						
Items	0 ^b	1	2	3	4	5	N
1. Adjusted Score:							
< 5	15.8	5.3	31.6	21.1	21.1	5.3	19
5 - 10	0.0	25.7	20.0	42.9	5.7	5.7	35
11 - 15	3.0	24.2	18.2	42.4	9.1	3.0	33
16 - 20	0.0	16.1	19.4	22.6	19.4	22.6	31
21 - 25	0.0	18.6	30.2	32.6	11.6	7.0	43
26 - 30	14.8	22.2	14.8	29.6	11.1	7.4	27
> 30	4.5	9.1	13.6	40.9	27.3	4.5	22
Missing							12
Total	4.3	18.6	21.4	33.8	13.8	8.1	222
2. Criterion-referenced test:							
< 5	5.3	0.0	15.8	42.1	21.1	15.8	19
5 - 10	0.0	2.9	14.7	32.4	17.6	32.4	34
11 - 15	0.0	12.5	3.1	10.6	25.0	18.8	32
16 - 20	3.3	3.3	0.0	30.0	36.7	26.7	30
21 - 25	2.3	9.3	4.7	30.2	30.2	23.3	43
26 - 30	3.7	18.5	11.1	25.9	22.2	18.5	27
> 30	4.5	9.1	9.1	31.8	22.7	22.7	22
Missing							15
Total	2.4	8.2	7.7	32.9	25.6	23.2	222
3. Grade equivalents (GE):							
< 5	15.0	5.0	5.0	15.0	35.0	25.0	20
5 - 10	0.0	2.9	11.4	17.1	22.9	45.7	35
11 - 15	3.0	3.0	3.0	24.2	24.2	42.4	33
16 - 20	3.3	3.3	0.0	13.3	26.7	53.3	30
21 - 25	0.0	2.3	4.7	18.6	39.5	34.9	43
26 - 30	0.0	3.7	7.4	22.2	29.6	37.0	27
> 30	0.0	9.1	9.1	13.6	18.2	50.0	22
Missing							12
Total	2.4	3.8	5.7	18.1	28.6	41.4	222

Table 17 (continued)

Percentage ^a of Responses							
Items	0 ^b	1	2	3	4	5	N
4. Mean:							
< 5	0.0	0.0	0.0	15.0	35.0	50.0	20
5 - 10	0.0	0.0	2.9	22.9	20.0	54.3	35
11 - 15	0.0	0.0	9.1	12.1	33.3	45.5	33
16 - 20	0.0	3.3	0.0	26.7	33.3	36.7	30
21 - 25	0.0	0.0	4.7	27.9	20.9	46.5	43
26 - 30	0.0	3.8	7.7	19.2	26.9	42.3	26
> 30	4.5	4.5	13.6	0.0	22.7	54.5	22
Missing							13
Total	0.5	1.4	5.3	19.1	26.8	46.9	222
5. Median:							
< 5	0.0	0.0	0.0	20.0	35.0	45.0	20
5 - 10	0.0	0.0	5.7	20.0	17.1	57.1	35
11 - 15	0.0	3.0	6.1	18.2	27.3	45.5	33
16 - 20	0.0	3.2	6.5	22.6	29.0	38.7	31
21 - 25	0.0	0.0	4.7	27.9	20.9	46.5	43
26 - 30	0.0	3.7	7.4	22.2	25.9	40.7	27
> 30	4.5	4.5	9.1	13.6	13.6	54.5	22
Missing							11
Total	0.5	1.9	5.7	21.3	23.7	46.9	222
6. Mode:							
< 5	0.0	0.0	5.0	25.0	25.0	45.0	20
5 - 10	0.0	0.0	11.4	22.9	20.0	45.7	35
11 - 15	0.0	9.1	9.1	18.2	24.2	39.4	33
16 - 20	0.0	9.7	6.5	25.8	22.6	35.5	31
21 - 25	0.0	4.8	7.1	31.0	19.0	38.1	42
26 - 30	0.0	3.7	18.5	14.8	37.0	25.9	27
> 300	4.5	9.1	4.5	13.6	27.3	40.9	22
Missing							12
Total	0.5	5.2	9.0	22.4	24.3	38.6	222

Table 17 (continued)

	Percentage ^a of Responses						
Items	0 ^b	1	2	3	4	5	N
7. Normal distribution:							
< 5	0.0	5.0	5.0	35.0	30.0	25.0	20
5 - 10	0.0	5.7	8.6	25.7	22.9	37.1	35
11 - 15	3.1	6.3	9.4	18.8	37.5	25.0	32
16 - 20	0.0	3.2	3.2	25.8	32.3	35.5	31
21 - 25	2.3	4.7	11.6	32.6	23.3	25.6	43
26 - 30	0.0	3.7	11.1	37.0	11.1	37.0	27
> 30	0.0	4.5	13.6	22.7	36.4	22.7	22
Missing							12
Total	1.0	4.8	9.0	28.1	27.1	30.0	222
8. Norm-referenced test:							
< 5	0.0	5.0	5.0	25.0	40.0	25.0	20
5 - 10	0.0	0.0	2.9	22.9	31.4	42.9	35
11 - 15	3.1	3.1	9.4	21.9	21.9	40.6	32
16 - 20	3.2	3.2	0.0	19.4	35.5	38.7	31
21 - 25	0.0	4.7	11.6	20.9	37.2	25.6	43
26 - 30	0.0	7.4	7.4	29.6	22.2	33.3	27
> 30	0.0	9.5	9.5	23.8	23.8	33.3	21
Missing							13
Total	1.0	4.3	6.7	23.0	30.6	34.4	222
9. Percentile:							
< 5	0.0	0.0	0.0	25.0	40.0	35.0	20
5 - 10	0.0	2.9	0.0	11.4	31.4	54.3	35
11 - 15	0.0	0.0	0.0	18.2	24.2	57.6	33
16 - 20	0.0	3.2	0.0	12.9	32.3	51.6	31
21 - 25	0.0	0.0	0.0	9.3	41.9	48.8	43
26 - 30	0.0	3.7	7.4	0.0	33.3	55.6	27
> 30	0.0	0.0	9.5	14.3	28.6	47.6	21
Missing							12
Total	0.0	1.4	1.9	12.4	33.3	51.0	222

Table 17 (continued)

Items	Percentage ^a of Responses						N
	0 ^b	1	2	3	4	5	
10. Percentile rank:							
< 5	0.0	0.0	5.0	20.0	40.0	35.0	20
5 - 10	0.0	2.9	2.9	17.1	22.9	54.3	35
11 - 15	0.0	3.0	0.0	15.2	24.2	57.6	33
16 - 20	0.0	3.2	0.0	16.1	25.8	54.8	31
21 - 25	0.0	0.0	0.0	11.6	37.2	51.2	43
26 - 30	0.0	3.7	7.4	0.0	33.3	55.6	27
> 30	0.0	0.0	13.6	18.2	22.7	45.5	22
Missing							11
Total	0.0	1.9	3.3	13.7	29.4	51.7	222
11. Raw score:							
< 5	0.0	0.0	15.0	40.0	10.0	35.0	20
5 - 10	0.0	2.9	5.7	20.0	31.4	40.0	35
11 - 15	0.0	12.1	3.0	15.2	18.2	51.5	33
16 - 20	0.0	3.2	3.2	25.8	22.6	45.2	31
21 - 25	0.0	0.0	2.3	25.6	27.9	44.2	43
26 - 30	0.0	3.7	3.7	18.5	25.9	48.1	27
> 30	0.0	4.5	9.1	18.2	18.2	50.0	22
Missing							11
Total	0.0	3.8	5.2	22.7	23.2	45.0	222
12. Reliability:							
< 5	0.0	5.0	15.0	25.0	40.0	15.0	20
5 - 10	2.9	0.0	14.3	20.0	34.3	28.6	35
11 - 15	6.1	12.1	12.1	36.4	18.2	15.2	33
16 - 20	0.0	12.9	9.7	29.0	25.8	22.6	31
21 - 25	0.0	4.7	9.3	37.2	25.6	23.3	43
26 - 30	0.0	7.4	11.1	7.4	37.0	37.0	27
> 30	4.5	4.5	18.2	31.8	13.6	27.3	22
Missing							11
Total	1.9	6.6	12.3	27.5	27.5	24.2	222

Table 17 (continued)

	Percentage ^a of Responses						
Items	0 ^b	1	2	3	4	5	N
13. Scale Score:							
< 5	5.0	10.0	35.0	15.0	25.0	10.0	20
5 - 10	2.9	2.9	17.1	22.9	28.6	25.7	35
11 - 15	6.1	12.1	15.2	33.3	21.2	12.1	33
16 - 20	6.5	16.1	12.9	22.6	19.4	22.6	31
21 - 25	0.0	4.7	11.6	41.9	20.9	20.9	43
26 - 30	0.0	7.7	15.4	19.2	11.5	46.2	26
> 30	4.5	13.6	18.2	13.6	40.9	9.1	22
Missing							12
Total	3.3	9.0	16.7	26.2	23.3	21.4	222
14. Standard Score:							
< 5	0.0	0.0	26.3	31.6	26.3	15.8	19
5 - 10	0.0	2.9	8.6	25.7	28.6	34.3	35
11 - 15	3.1	9.4	9.4	28.1	31.3	18.8	32
16 - 20	3.2	6.5	16.1	22.6	19.4	32.3	31
21 - 25	0.0	2.3	11.6	30.2	30.2	25.6	43
26 - 30	0.0	7.4	7.4	29.6	11.1	44.4	27
> 30	4.5	9.1	22.7	9.1	40.9	13.6	22
Missing							13
Total	1.4	5.3	13.4	25.8	26.8	27.3	222
15. Stanine:							
< 5	0.0	10.0	20.0	25.0	40.0	5.0	20
5 - 10	5.9	0.0	2.9	17.6	35.3	38.2	34
11 - 15	3.0	12.1	0.0	15.2	27.3	42.4	33
16 - 20	0.0	3.2	9.7	19.4	25.8	41.9	31
21 - 25	0.0	0.0	4.7	20.9	37.2	37.2	43
26 - 30	0.0	11.1	0.0	7.4	29.6	51.9	27
> 30	0.0	4.5	13.6	13.6	27.3	40.9	22
Missing							12
Total	1.4	5.2	6.2	17.1	31.9	38.1	222

Table 17 (continued)

Percentage ^a of Responses							
Items	0 ^b	1	2	3	4	5	N
16. Statistical process control:							
< 5	10.0	45.0	15.0	15.0	10.0	5.0	20
5 - 10	8.6	31.4	20.0	17.1	5.7	17.1	35
11 - 15	15.2	33.3	21.2	18.2	9.1	3.0	33
16 - 20	9.7	32.3	12.9	22.6	16.1	6.5	31
21 - 25	4.8	28.6	23.8	33.3	7.1	2.4	42
26 - 30	11.1	25.9	18.5	22.2	0.0	22.2	27
> 30	9.1	27.3	22.7	22.7	13.6	4.5	22
Missing							12
Total	9.5	31.4	19.5	22.4	8.6	8.6	222
17. Validity:							
< 5	0.0	5.0	25.0	30.0	15.0	25.0	20
5 - 10	2.9	2.9	17.1	17.1	28.6	31.4	35
11 - 15	6.1	21.2	12.1	30.3	12.1	18.2	33
16 - 20	0.0	22.6	12.9	12.9	19.4	32.3	31
21 - 25	2.3	4.7	20.9	25.6	30.2	16.3	43
26 - 30	0.0	14.8	11.1	18.5	25.9	29.6	27
> 30	0.0	9.1	13.6	18.2	50.0	9.1	22
Missing							11
Total	1.9	11.4	16.1	21.8	25.6	23.2	222

^aPercentages may not equal 100% due to rounding

^bResponses were from 1= Not Well to 3 = Fairly Well to 5 = Very Well, with 0 = Don't Know

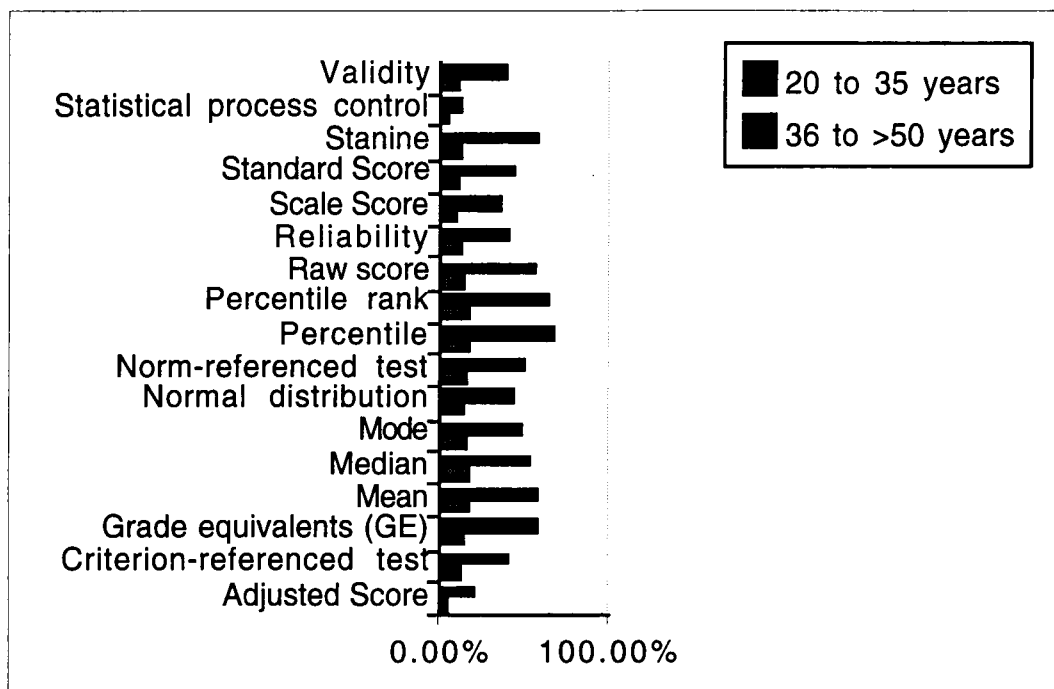


Figure 1. A Comparison of Term Perception vs. Age

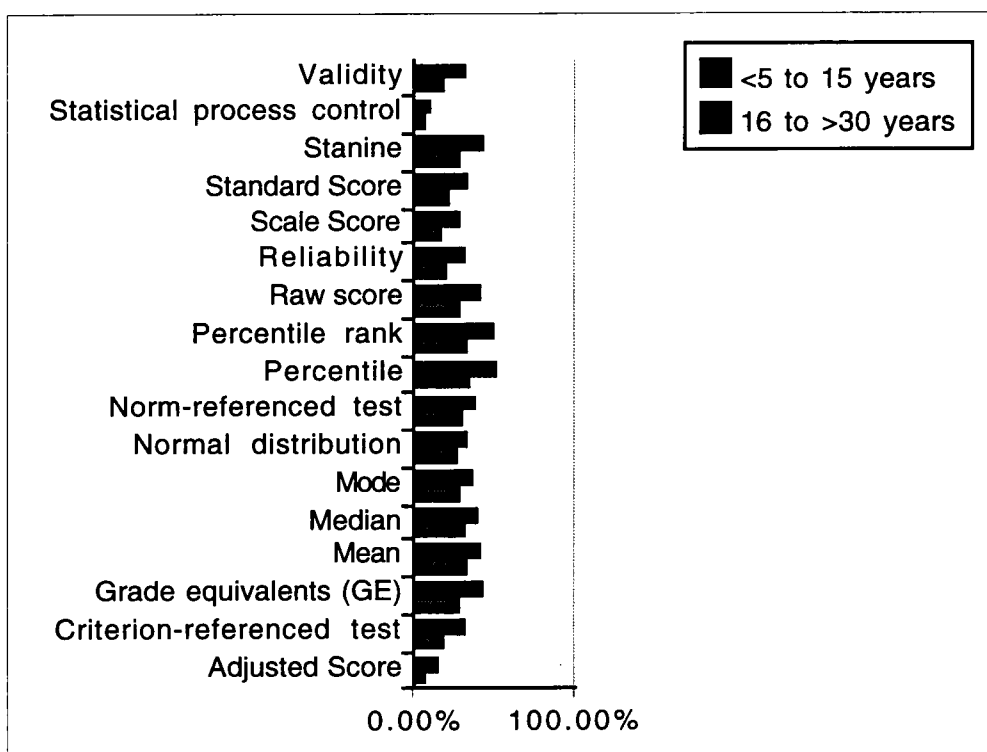


Figure 2. A Comparison of Term Perception vs. Years' Experience

percentages of those respondents who indicated they could explain a term very well or slightly less (a ranking of 4 or 5) between the groups of from 0 to 15 years' experience and those with greater than 16 years' experience. As can be seen, the percentages are not as large as when compared with the respondents' ages. Still, in every instance, the percentage of those who stated they could explain a term to another is greater as the years' experience increased.

While there was an increase in the number who stated they would feel confident defining a term for another, the increase was not as great as it was when compared to the respondents' age. For example, the term normal distribution, saw an increase in the 41 to 45 year old age group, but not such a large increase in those with between 21 and 25 years' experience, the corresponding group. Such a deficiency could have been caused by the assignments of the respondents or a limitation of exposure to terms normally associated with tests and measurement.

The sample population was asked in what context they used the terms listed in the previous item. The respondents were given a variety of situations and asked in which they used the terms. The respondents could mark more than one situation. The results are shown in Table 18. Those respondents who completed Item 8, Other, supplied responses that varied from "Curriculum placement," "In-Service," and "Staff development," to "Setting goals," "Planning/Scheduling," and "Talking with other teachers," with In-Service being the most mentioned response. The most frequent use of the terms listed in the previous items appeared in Multi-Disciplinary Team (M-Team) Meetings. This is not surprising, since those type of meetings are normally held to discuss a student's placement and instructional program based on scores of various tests, including Intelligence Quotient (IQ) and other standardized tests. The second most frequent use appeared in Parent-Teacher conferences, which again is not surprising because test scores are often mentioned in conjunction with a student's progress--or lack of it.

Table 18

Contexts of Term Use

When term used	Responses	
	Frequency	%
Multi-Disciplinary (M-Team) Meetings	150	67.6
Parent-Teacher Conferences	147	66.2
Faculty Meetings	131	59.0
Workshop/Seminars	115	51.8
Teacher Evaluation Conferences	70	31.5
Central Office Meetings	64	28.8
Student Conferences	50	22.5
Other	22	9.9

Because the M-Team Meeting was the most prevalent context of term usage, what was the most-used term in that context? Table 19 shows the use of the previous item terms in the context of an M-Team Meeting. The most-frequently used term in M-Team meetings was percentile, with respondents stating that term was used 80.4% of the time (from "Sometimes Use" to "Frequently Use"). Percentile rank was the next most-used term with 77.3% indicating they used that term. Stanine was third-ranked with 68.3% stating it was used, followed closely by Grade Equivalents, 64.4%. Statistical Process Control was the least-used term, with only 12.8% of the respondents indicating they used that term "Sometimes" to "Frequently."

Previous Training in Psychometrics

Another portion of the study dealt with the training received by educators in the state of Tennessee. To attempt to determine the level of professional training regarding tests and measurement, respondents were asked if they had ever taken a specific course in tests and measurement. Their responses were limited to indicating if they had taken a complete course, a partial course with some aspects of testing, never had taken a course, or that they didn't remember if they had taken a course or not. Of the respondents in the survey, 121 stated they had taken a complete course in tests and measurement, 43 stated they had taken at least a partial course, 38 had never taken such a course, 10 did not remember, and 10 left the question blank. Consequently, 73.3% of those in the population who responded to the question had taken such a course or at least a partial course in test and measurements.

Those who had indicated they had taken a full or a partial course in tests and measurement were asked to describe when and where, and in which department, that they had taken that course. Respondents could mark more than one response. The results of that question are shown in Table 20. Those who indicated "Other" on the survey

Table 19

Percentage of Use of Testing Terms in M-Team Meetings

Terms	Responses ^a					N ^c
	1 ^b	2	3	4	5	
1. Adjusted Score	61.1	20.1	12.8	4.7	1.3	149
2. Criterion-referenced test	31.8	20.3	33.1	8.8	6.1	148
3. Grade equivalents (GE)	18.1	17.4	33.6	17.4	13.4	149
4. Mean	20.8	26.2	33.6	10.1	9.4	149
5. Median	27.0	23.0	35.1	8.8	6.1	148
6. Mode	36.2	22.1	28.9	8.7	4.0	149
7. Normal distribution	38.5	21.6	27.0	8.1	4.7	148
8. Norm-referenced test	29.1	25.0	27.7	10.8	7.4	148
9. Percentile	10.1	9.5	35.1	22.3	23.0	148
10. Percentile rank	12.8	10.1	33.6	24.2	19.5	149
11. Raw score	25.0	26.4	29.1	13.5	6.1	148
12. Reliability	32.4	21.6	24.3	16.9	4.7	148
13. Scale Score	36.5	21.6	29.1	9.5	3.4	148
14. Standard Score	27.7	22.3	31.8	11.5	6.8	148
15. Stanine	16.9	14.9	27.7	23.0	17.6	148

Table 19 (continued)

Terms	Responses ^a					N ^c
	1 ^b	2	3	4	5	
16. Statistical process control	72.5	14.8	10.1	2.7	0.0	149
17. Validity	36.2	22.8	24.2	11.4	5.4	149

^aResponses were labeled 1= Never Use 3 = Sometimes Use 5 = Frequently Use

^bPercentages may not add to 100% due to rounding

^cNumber responding (of the 222 participants)

Table 20

Course Descriptions

When/Where course taken	Responses	
	Number of Responses	% ^a
1. Undergraduate level	108	48.6
2. Graduate level	98	44.1
3. Business Department	5	2.3
4. Curriculum and Instruction Department	84	37.8
5. Educational Administration and Supervision Department	34	15.4
6. Educational Psychology Department	35	15.8
7. Liberal Arts Department	3	1.4
8. Other	13	5.9

^aPercentage is greater than 100% due to multiple responses

instrument gave responses from "Elementary Education," "Special Education," "TAE (with no explanation)," "Workshops," to "tests and measurement in HPED," with courses in Special Education being the most-received response. According to the respondents, the majority of courses taken were at the undergraduate level in the Department of Curriculum and Instruction.

In an attempt to ascertain the usefulness of those courses, respondents, if they had taken a course in tests and measurement, were asked how frequently they had utilized material from that course. Given a scale from "Not very Frequently" to "Very Frequently," the responses indicated that the majority, 49.1%, did not use any material learned from a course that frequently. Only 3.2% stated they did not use material very frequently, and 10% stated they used material frequently or very frequently. The information about tests and measurement gained through college courses did not appear to be used with great regularity by those educators responding to the survey.

Along the same lines, respondents were asked if they had ever undertaken work independently to increase their level of knowledge of tests and measurement. The rationale here was that if educators were involved enough in their work and did not have access to courses or material about the subject, that they would voluntarily work on increasing their level of information on the subject. Given the choices of "Yes," "No," or "Don't Remember," 37.4% said yes, 57.7% said no, 2.7% stated they could not remember, and 2.3% left the question blank. Of those who answered "Yes" to the previous question, they were asked what method they used to increase their knowledge. Their responses are shown in Table 21. The use of books or tapes and the use of journals or articles headed the responses with 45 and 44 instances respectively. The next most-frequently used method was discussions with others teachers, guidance counselors, or principals. Knowledge gained at testing seminars or workshops was fourth. The high incidence of independent study pointed out a need for better and more comprehensive instruction of educators in the area of tests and measurement. While a majority said

Table 21

Use of Outside Activities to Increase Knowledge of Testing

Outside Activity	Responses
	Frequency
Personal study of books and tapes	45
Use of journals/articles	44
Discussions with teachers/guidance/principal/others	15
Classes	13
Testing seminars/workshops	12
Use of test manuals/trade information	9
Faculty meetings/in-service training	8
Material sent from central office/bulletins	4
Studying for career ladder evaluation	4
Reviews of tests/updates of test versions	3
Curriculum from other courses	2
Field testing/administration of testing instruments	2
Computer training	1

Note. Responses were to an open-ended question; some responses were combined into one category.

they did not participate in independent study, the perceptions of educators was low, with term use even lower.

Respondents were asked how frequently they had utilized the technical manual that accompanied each standardized test. Respondents were not asked if they were aware a technical manual accompanied each test, only if they had used it. This was an attempt to determine if not only did the respondent know a technical manual accompanied a standardized test, but also if it were examined by any of the respondents. The mean response was 2.95, almost exactly halfway between "Never" and "Often." A large majority, 62.2%, stated that they never or almost never used the manual. One respondent stated that he had never even seen one, especially the manual that accompanied the T-CAP test.

To determine on-going efforts at staff development in the area of tests and measurement by schools and districts, respondents were asked how often during the year their school or school district engaged in activities with teachers related to testing. The results are shown in Table 22. Since such on-going efforts of both schools or districts to inform and train their teachers are an important part of maintaining understanding and level of knowledge of their staff, such efforts would show an awareness of a possible area of weakness and the efforts to strengthen information in that area (Marso & Pigge, 1992). Efforts by schools/districts are deemed important if they address specific competencies directly related to improving teachers' methods and practices in the area of testing and use of testing.

Judging by the means of the scores received, schools/districts make some efforts to improve test scores and to strengthen instruction in specific areas where test scores show weakness. While such endeavors are directed toward improving test scores, the schools/districts seem to make less attempts to assist teachers in specific testing and evaluation techniques. This could be a factor leading to the lack of knowledge demonstrated by the responses to the earlier questions concerning testing knowledge and

Table 22

Frequency of Activities Related to Testing

Items	Responses	
	N ^a	Mean ^b
Discusses methods to strengthen instruction in the specific areas where test scores show weakness	215	3.00
Discusses methods to improve test scores	215	2.96
Holds staff meetings to review test scores	214	2.66
Holds in-service specifically devoted to tests and measurements	213	2.37
Assists teachers in specific testing and evaluation techniques	214	2.33

^aNumber responding (of the 222 participants)

^bResponses were from 1= Never to 5 = Often

score use.

The final items on the survey instrument attempted to determine if any respondents had taught earlier in a state or district where student standardized test results were used as a component of teacher evaluation. Of the eight "Yes" responses that were received to the question, none were valid because of the incompleteness of the replies. This might have been due to the wording of the item. When prompted for a reply, the respondent was to also indicate where this method of evaluation had occurred. Only one respondent indicated "Yes" and then inserted "TN" for a location with no amplifying information. The remaining 214 responses were "No." Because there were no valid responses to the item, there were no valid responses to the remainder of the items dealing with thoughts about the use of student test scores for teacher evaluation. Unfortunately, the lack of valid responses precluded an analysis of those attitudes of experienced personnel in this area. Perhaps future research could rectify this lack of information.

Attitudes toward Standardized Tests, "Value-Added" **Assessment, and Changes in Teaching Methods**

In an attempt to determine attitudes towards standardized tests in general, respondents were asked the degree that they agreed or disagreed with statements describing standardized tests. The results are shown in Table 23. From the results, it would appear that the educators surveyed agreed somewhat with the statements that standardized tests tested lower-order thinking skills in the cognitive domain. They were not unanimous in their agreement, however. The responses for the items tended to cluster around the mid point, which is indicative of a neutral response, neither agreeing or disagreeing with the statements.

The most frequent responses tended to be low to mid-range, as stated above, with

Table 23

Attitudes Concerning Standardized Testing

Items	Responses	
	N ^a	Mean ^b
Test the cognitive domain	210	3.02
Test lower-order thinking skills	210	2.99
Test higher-order thinking skills	210	2.57
Reliable method for diagnosing student problems	211	2.50
Valid means of assessing student achievement	212	2.42
Reliable means of assessing student achievement	213	2.40
Test the affective domain	209	2.22

^aNumber responding (of the 222 participants)

^bResponses were from 1= Strongly Disagree to 5 = Strongly Agree

the exception of the scores for those items related to testing lower-order cognitive skills. This is in keeping with other findings that support that teachers tend to believe that standardized tests test only lower-order skills and only information that can be easily recalled (Bracey, 1992; Dorr-Bremme & Herman, 1986; Johns & Davis, 1991; Marso & Pigge, 1992). As far as higher-order thinking skills and the affective domain are concerned, it has been suggested that they are better assessed by other methods (Beatty, 1969; Bowers, 1989).

This can be seen more clearly when the age of the respondents was compared with those same attitudes toward standardized tests as shown in Table 24. The respondents gave a neutral to low response to the item that dealt with providing a reliable means of assessing student achievement. Among all groups, there was a swing toward the lower end of the scale, with the majority in many cases in the middle. The second item produced a result that indicated the younger group, those younger than 41 years old, felt that standardized tests did not provide a valid means of assessing student achievement. The older group, those older than 41, were more neutral in their attitudes toward this item with majorities in the middle area--neither agreeing nor disagreeing. The results to the third item, that dealing with providing a reliable method of diagnosing student problems, provided a similar pattern. Those younger than 31 did not agree that standardized tests were a reliable method of diagnosing student problems. Those older than 31 were somewhat higher in their scores, yet were still neutral in their attitudes toward standardized test for this purpose except for the 41 to 45 age group, who sided with their younger colleagues.

As a whole, the respondents felt more strongly that standardized tests tested the cognitive domain, which was not unexpected. While the majorities indicated a neutral stance to the item with scores of three, a greater number of higher scores, indicating agreement, were recorded. The oldest group, those above 50 years old, were in tandem with the younger respondents, with the majority also indicating that they felt neutral

Table 24

Attitudes toward Standardized Tests by Age

	Percentage ^a of Responses					
Items	1 ^b	2	3	4	5	N
1. Reliable assessment						
20 - 25	0.0	75.0	25.0	0.0	0.0	4
26 - 30	17.2	34.5	41.4	6.9	0.0	29
31 - 35	16.7	50.0	33.3	0.0	0.0	12
36 - 40	40.0	32.0	28.0	0.0	0.0	25
41 - 45	25.5	34.5	38.2	1.8	0.0	55
46 - 50	8.3	19.4	61.1	8.3	2.8	36
> 50	13.5	23.1	44.2	11.5	7.7	52
Missing						9
Total	19.5	30.5	42.3	5.6	2.3	222
2. Valid assessment						
20 - 25	0.0	75.0	25.0	0.0	0.0	4
26 - 30	13.8	48.3	37.9	0.0	0.0	29
31 - 35	8.3	66.7	25.0	0.0	0.0	12
36 - 40	40.0	36.0	24.0	0.0	0.0	25
41 - 45	18.5	38.9	38.9	3.7	0.0	54
46 - 50	13.9	19.4	56.5	5.6	2.8	36
> 50	9.6	25.0	36.5	21.2	7.7	52
Missing						10
Total	16.5	35.4	38.7	7.1	2.4	222
3. Diagnosing problems						
20 - 25	0.0	100.0	0.0	0.0	0.0	4
26 - 30	13.8	41.4	34.5	6.9	3.4	29
31 - 35	33.3	16.7	50.0	0.0	0.0	12
36 - 40	28.0	28.0	28.0	12.0	4.0	25
41 - 45	18.5	40.7	35.2	5.6	0.0	54
46 - 50	11.4	22.9	48.6	17.1	0.0	35
> 50	15.4	26.9	26.9	23.1	7.7	52
Missing						11
Total	17.5	32.7	34.6	12.3	2.8	222

Table 24 (continued)

	Percentage ^a of Responses					
Items	1 ^b	2	3	4	5	N
4. Tests cognitive domain						
20 - 25	0.0	0.0	50.0	50.0	0.0	4
26 - 30	10.3	24.1	37.9	24.1	3.4	29
31 - 35	0.0	33.3	25.0	33.3	8.3	12
36 - 40	16.0	20.0	48.0	8.0	8.0	25
41 - 45	11.3	24.5	28.3	28.3	7.5	53
46 - 50	0.0	16.7	47.2	33.3	2.8	36
> 50	11.8	11.8	43.1	19.6	13.7	51
Missing						12
Total	9.0	19.5	39.0	24.8	7.6	222
5. Tests affective domain						
20 - 25	50.0	0.0	50.0	0.0	0.0	4
26 - 30	17.2	31.0	44.8	3.4	3.4	29
31 - 35	41.7	33.3	25.0	0.0	0.0	12
36 - 40	29.2	37.5	29.2	4.2	0.0	24
41 - 45	41.5	34.0	22.6	0.0	1.9	53
46 - 50	16.7	41.7	33.3	5.6	2.8	36
> 50	23.5	19.6	39.2	13.7	3.9	51
Missing						13
Total	28.2	31.1	33.0	5.3	2.4	222
6. Tests higher-order skills						
20 - 25	50.0	0.0	25.0	25.0	0.0	4
26 - 30	10.3	27.6	41.4	17.2	3.4	29
31 - 35	25.0	16.7	41.7	16.7	0.0	12
36 - 40	20.8	37.5	29.2	8.3	4.2	24
41 - 45	34.0	20.8	35.8	9.4	0.0	53
46 - 50	5.6	30.6	30.6	30.6	2.8	36
> 50	19.2	23.1	32.7	19.2	5.8	52
Missing						12
Total	20.5	25.2	34.3	17.1	2.9	222

Table 24 (continued)

Items	Percentage ^a of Responses					N
	1 ^b	2	3	4	5	
7. Tests lower-order skills						
20 - 25	0.0	25.0	50.0	25.0	0.0	4
26 - 30	6.9	27.6	34.5	24.1	6.9	29
31 - 35	8.3	16.7	41.7	0.0	33.3	12
36 - 40	33.3	16.7	29.2	16.7	4.2	24
41 - 45	15.1	20.8	26.4	24.5	13.2	53
46 - 50	2.8	16.7	33.3	41.7	5.6	36
> 50	17.3	9.6	42.3	21.2	9.6	52
Missing						12
Total	13.8	17.6	34.3	24.3	10.0	222

^aPercentages may not add to 100% due to rounding

^bResponses were from 1= Strongly Disagree to 5 = Strongly Agree

toward this item. When asked about testing the affective domain, the age groups were all in agreement that standardized tests did not test the affective domain. Once again the majority in the oldest group indicated a neutral stance, neither agreeing nor disagreeing with the statement.

There was agreement among all age groups as to whether standardized tests tested higher-order thinking skills or not. For the most part, they indicated that they disagreed with that statement. The age group of those between 46 and 50 were more neutral in their attitudes, with equal scores for two, three, and four. The final statement in this item deal with testing lower-order thinking skills. All groups leaned more toward agreement than disagreement. The older group once again displayed its neutral stance with the majority of those over 50 neither agreeing nor disagreeing. The 36 to 40 age group dissented from the remainder, however, with the majority strongly disagreeing with the statement. The relatively neutral answers of all groups toward these statements indicates no prevalent attitudes toward standardized tests in the abstract. Taken in conjunction with the earlier analysis, there appears to be a consensus about the usefulness of standardized tests as reliable, valid measures of student achievement or as a means of diagnosing student problems.

When respondents' years of experience were compared with their attitudes toward standardized tests, the same type of responses were found. The results can be found in Table 25. Most responses were at the mid-range point, indicating neither agreement nor disagreement with the statements. The first statement, that dealing with reliability of standardized tests for student achievement assessment, found a neutral answer from the majority of all groups except the ones which had the fewest years of experience and the 11 to 15 years' experience group. Those groups tended more toward disagreement than agreement. Much the same attitudes were expressed by the same groups concerning the next statement, the one regarding the validity of standardized tests for assessing student achievement. The respondents with the least number of years'

Table 25

Attitudes toward Standardized Tests by Years' Experience

Items	Percentage ^a of Responses					N
	1 ^b	2	3	4	5	
1. Reliable assessment						
< 5	10.0	45.0	40.0	5.0	0.0	20
5 - 10	22.9	34.3	40.0	2.9	0.0	35
11 - 15	30.3	42.4	24.2	3.0	0.0	33
16 - 20	19.4	29.0	45.2	3.2	3.2	31
21 - 25	16.3	23.3	51.7	4.7	2.3	43
26 - 30	11.1	22.2	51.9	11.1	3.7	27
> 30	18.2	18.2	40.9	13.6	9.1	22
Missing						11
Total	19.0	30.3	42.7	5.7	2.4	222
2. Valid assessment						
< 5	10.0	50.0	40.0	0.0	0.0	20
5 - 10	20.0	45.7	31.4	2.9	0.0	35
11 - 15	25.0	53.1	15.6	6.3	0.0	32
16 - 20	19.4	35.5	38.7	6.5	0.0	31
21 - 25	16.3	23.3	51.7	4.7	2.3	43
26 - 30	7.4	25.9	48.1	14.8	3.7	27
> 30	9.1	18.2	40.9	18.2	13.6	22
Missing						12
Total	16.2	35.7	38.6	7.1	2.4	222
3. Diagnosing problems						
< 5	10.0	50.0	35.0	5.0	0.0	20
5 - 10	17.1	40.0	37.1	2.9	2.9	35
11 - 15	21.9	34.4	28.1	15.6	0.0	32
16 - 20	13.3	30.0	43.3	10.0	3.3	30
21 - 25	25.6	25.6	32.6	14.0	2.3	43
26 - 30	14.8	25.9	37.0	14.8	7.4	27
> 30	9.1	27.3	31.8	27.3	4.5	22
Missing						13
Total	17.2	32.5	34.9	12.4	2.9	222

Table 25 (continued)

	Percentage ^a of Responses					
Items	1 ^b	2	3	4	5	N
4. Tests cognitive domain						
< 5	10.0	20.0	45.0	25.0	0.0	20
5 - 10	8.6	28.6	31.4	22.9	8.6	35
11 - 15	15.6	21.9	28.1	28.1	6.3	32
16 - 20	6.5	12.9	21.6	22.6	6.5	31
21 - 25	4.8	26.2	35.7	26.2	7.1	42
26 - 30	11.1	11.1	40.7	29.6	7.4	27
> 30	9.5	9.5	47.6	19.0	14.3	21
Missing						14
Total	9.1	19.7	38.9	25.0	7.2	222
5. Tests affective domain						
< 5	25.0	25.0	20.0	0.0	0.0	20
5 - 10	28.6	37.1	25.7	2.9	5.7	35
11 - 15	37.5	37.5	21.9	3.1	0.0	32
16 - 20	45.2	29.0	25.8	0.0	0.0	31
21 - 25	26.2	38.1	35.7	0.0	0.0	42
26 - 30	14.8	22.2	44.4	11.1	7.4	27
> 30	14.3	19.0	38.1	28.6	0.0	21
Missing						14
Total	28.4	31.3	33.2	5.3	1.9	222
6. Tests higher-order skills						
< 5	30.0	10.0	40.0	20.0	0.0	20
5 - 10	11.4	37.1	37.1	8.6	5.7	35
11 - 15	25.0	40.6	25.0	9.4	0.0	32
16 - 20	29.0	22.6	38.7	9.7	0.0	31
21 - 25	16.7	23.8	26.2	33.3	0.0	42
26 - 30	14.8	11.1	44.4	22.2	7.4	27
> 30	22.7	18.2	36.4	13.6	9.1	22
Missing						13
Total	20.6	24.9	34.4	17.2	2.9	222

Table 25 (continued)

Items	Percentage ^a of Responses					N
	1 ^b	2	3	4	5	
7. Tests lower-order skills						
< 5	10.0	15.0	35.0	25.0	15.0	20
5 - 10	8.6	37.1	25.7	17.1	11.4	35
11 - 15	21.9	9.4	31.3	28.1	9.4	32
16 - 20	9.7	12.9	35.5	29.0	12.9	31
21 - 25	11.9	19.0	28.6	31.0	9.5	42
26 - 30	11.1	11.1	48.1	22.2	7.4	27
> 30	27.3	9.1	45.5	13.6	4.5	22
Missing						13
Total	13.9	17.2	34.4	24.4	10.0	222

^aPercentages may not add to 100% due to rounding

^bResponses were from 1= Strongly Disagree to 5 = Strongly Agree

experience, less than 15 years, indicated disagreement with the statement, while the ones with more experience displayed a more neutral attitude to this statement. The third statement in this item, the one regarding the reliability of standardized tests for diagnosing student problems, elicited much the same response. The group with fewer years' experience, less than 15 years', indicated disagreement with the statement. The remaining groups indicated the more neutral response. Over all, the responses tended more to the negative, disagreeing with the statement.

For the statements concerning the cognitive and affective domains, the educators with between 11 and 25 years' experience tended to agree more with the statement about testing the cognitive domain and tended to disagree strongly with the statement regarding affective domain testing. The groups with the least experience and the most experience remained neutral on testing both the cognitive and affective domains. When statements about testing higher- and lower-order thinking skills were posed, the attitudes of the respondents reflected a tendency to disagree with the statement concerning the testing of higher-order skills, but showed more agreement concerning the testing of lower-order skills. As stated above, this was in keeping with other's findings regarding the attitudes of educators toward the use of standardized tests for testing lower-order thinking skills. The educators in Tennessee were no exception. The respondents with the greater years of experience indicated agreement with the statement that standardized tests do test lower-order thinking skills much better than they test higher-order skills with the majority in each group tending toward agreement. For all statements in this item, there was little strong agreement. Most of the respondents, especially those with over 30 years' experience, tended to be more neutral on the statements by neither agreeing nor disagreeing. This may be an indication of uncertainty as to the effectiveness of standardized tests and exactly what they test, or it may be a part of reticence on the respondents to take a stand on this issue.

Next, respondents were asked their attitudes toward the new requirements by the

state of Tennessee regarding the use of student test results in its "value-added" assessment program. In this requirement, "gain" scores are to be computed for each student based on his or her beginning level of performance and his or her ending level of performance (Cook, 1984; McLean & Sanders, 1984, 1991). This score will then become a part of the teacher's performance evaluation for that year. Participants were asked to indicate their agreement or disagreement with statements related to the use of student test scores for teacher evaluation. The results are shown in Table 26.

This survey question received particularly low scores indicating the attitude of surveyed educators toward the state's new evaluation plan. Given the attitude of educators toward the perceived effectiveness of standardized tests, it was not surprising that they would not desire their performance evaluation tied to results of those tests. As can be seen, low scores on reliability, validity, and accuracy were in keeping with previously displayed attitudes toward tests in general. Also, the low score on the item concerned with teacher evaluation tends to support the other scores as well. A majority of 76.1% either strongly disagreed or disagreed with the statement that standardized tests should be part of a teacher's evaluation; only 2.7% strongly agreed with that statement. The largest agreement was on the statement that concerned the reliability of a standardized test as a means of assessing teacher performance. Here, 84.3% of the educators strongly disagreed or disagreed that standardized tests can reliably assess a teacher's performance. While this statement received the majority of responses, its mean was the lowest at 1.50. The statement with the highest mean, the one concerned with tests reflecting the student's achievement gains, found 33.8% strongly agreeing with it.

The respondents' age was compared with attitudes toward the use of student standardized test data in teacher evaluation. Respondents were asked if they agreed or disagreed with the statement; extremes were labeled "strongly disagree" and "strongly agree." The results are shown in Table 27. The first statement in this item was

Table 26

Attitudes Concerning Using "Value-Added" Test Scores as a Component of Teacher Evaluation

Items ^a	Responses	
	N ^b	Mean ^c
Reflect students' achievement gains	212	2.16
Can indicate teacher effects on students' achievements	212	1.98
Represent an accurate picture of the breadth of student's learning	212	1.72
Should be a component of teacher evaluation	211	1.68
Are a valid means of assessing teacher performance	212	1.52
Are a reliable means of assessing teacher performance	212	1.50

^aItems were statements beginning with "Standardized test results. . ."

^bNumber responding (of the 222 participants)

^cResponses were from 1= Strongly Disagree to 5 = Strongly Agree

Table 27

Attitudes toward "Value-Added" Assessment by Age

	Percentage ^a of Responses					
Item ^b	1 ^c	2	3	4	5	N
1. Should be a component of teacher evaluation:						
20 - 25	25.0	50.0	25.0	0.0	0.0	4
26 - 30	62.1	17.2	17.2	3.4	0.0	29
31 - 35	75.0	8.3	16.7	0.0	0.0	12
36 - 40	84.0	16.0	0.0	0.0	0.0	25
41 - 45	64.8	22.2	11.1	0.0	1.9	54
46 - 50	47.2	27.8	22.2	2.8	0.0	36
> 50	49.1	19.6	13.7	7.8	9.8	51
Missing						11
Total	59.7	20.9	13.7	2.8	2.8	222
2. Are a valid means of assessing teacher performance:						
20 - 25	25.0	75.0	0.0	0.0	0.0	4
26 - 30	69.0	20.7	6.9	3.4	0.0	29
31 - 35	83.3	16.7	0.0	0.0	0.0	12
36 - 40	80.0	16.0	0.0	4.0	0.0	25
41 - 45	70.4	20.4	9.3	0.0	0.0	54
46 - 50	61.1	19.4	16.7	2.8	0.0	36
> 50	51.9	28.8	9.6	3.8	5.8	52
Missing						10
Total	65.1	22.6	8.5	2.4	1.4	222
3. Are a reliable means of assessing teacher performance:						
20 - 25	25.0	75.0	0.0	0.0	0.0	4
26 - 30	65.5	24.1	10.3	0.0	0.0	29
31 - 35	75.0	25.0	0.0	0.0	0.0	12
36 - 40	88.0	8.0	0.0	4.0	0.0	25
41 - 45	72.2	18.5	9.3	0.0	0.0	54
46 - 50	66.7	22.2	11.1	0.0	0.0	36
> 50	48.1	30.8	9.6	5.8	5.8	52
Missing						10
Total	65.6	23.1	8.0	1.9	1.4	222

Table 27 (continued)

	Percentage ^a of Responses					
Item ^b	1 ^c	2	3	4	5	N
4. Can indicate teacher effects on students' achievements:						
20 - 25	25.0	50.0	25.0	0.0	0.0	4
26 - 30	41.4	24.1	24.1	10.3	0.0	29
31 - 35	41.7	33.3	16.7	8.3	0.0	12
36 - 40	72.0	12.0	12.0	4.0	0.0	25
41 - 45	53.7	25.9	16.7	3.7	0.0	54
46 - 50	41.7	16.7	27.8	13.9	0.0	36
Missing						10
Total	45.3	24.1	19.3	9.9	1.4	222
5. Represent an accurate picture of the breadth of students' learning:						
20 - 25	25.0	75.0	0.0	0.0	0.0	4
26 - 30	58.6	17.2	20.7	3.4	0.0	29
31 - 35	66.7	16.7	16.7	0.0	0.0	12
36 - 40	84.0	8.0	4.0	4.0	0.0	25
41 - 45	59.3	24.1	13.0	3.7	0.0	54
46 - 50	44.4	30.6	19.4	2.8	2.8	36
> 50	38.4	34.6	17.3	5.8	3.8	52
Missing						10
Total	54.3	25.5	15.1	3.8	1.4	222

Table 27 (continued)

	Percentage ^a of Responses					
Item ^b	1 ^c	2	3	4	5	N
6. Reflect students' achievement gains:						
20 - 25	25.0	25.0	50.0	0.0	0.0	4
26 - 30	31.0	37.9	27.6	3.4	0.0	29
31 - 35	41.7	33.3	16.7	8.3	0.0	12
36 - 40	72.0	12.0	16.0	0.0	0.0	25
41 - 45	37.0	33.3	27.8	0.0	1.9	54
46 - 50	36.1	16.7	33.3	13.9	0.0	36
> 50	17.3	26.9	30.8	13.5	11.5	52
Missing						10
Total	35.4	26.9	27.8	6.6	3.3	222

^aPercentages may not add to 100% due to rounding

^bItems were statements beginning with "Standardized test results. . ."

^cResponses were from 1= Strongly Disagree to 5 = Strongly Agree

concerned with whether standardized test results should be a component of teacher evaluation or not. The majority in all age groups except those from 20 to 25, strongly disagreed with the statement, with the largest percentage coming from the 31 to 35 and the 36 to 40 age groups, 75% and 84% respectively. This was a strong indication about how educators felt about the new evaluation requirements.

The next statement asked if the respondents agreed or disagreed that standardized test results were a valid means of assessing teacher performance. Again, all age groups disagreed strongly with the statement, with the same two age groups, the 31 to 35 and the 36 to 40 age groups, having the largest percentages who strongly disagreed, 100% and 88.9% respectively. The third statement in this item was concerned with whether or not standardized test results were a reliable means of assessing teacher performance. Paralleling the earlier results, the majority of all age groups disagreed strongly with this statement. These findings indicate that educators in Tennessee do not feel that standardized tests are a valid or reliable means of assessing teacher performance, and they should not be a component of teacher evaluations.

The fourth statement in this item dealt with standardized test results being an indication of teacher effects on students' achievements. Could a standardized test measure the effectiveness of a teacher on a student's accomplishments? While the respondents disagreed strongly with this statement, those in the older age groups were not as universal in their disagreement. As the age increased above 41, the number of those strongly disagreeing decreased. In the 46 to 50 year old age group, almost as many strongly disagreed as felt neutral or slightly less than neutral on the subject. With age, apparently, there has evolved more understanding.

The last two statements in this item were concerned with whether or not standardized test results represented an accurate picture of the breadth of students' learning and reflected students' achievement gains. For the former, the a majority of the respondents in all age groups except the youngest group strongly disagreed with the

statement. For the latter statement, the disagreement was not as strong. The 20 to 30 age groups, the 41 to 45 age group, and the greater than 50 years old age group indicated they were not as strong in their disagreement as the other groups, with the majority tending more toward a neutral position. In fact, in the greater than 50 age group, the majority was neutral, with 30.8% reflecting that opinion. These responses indicate an overall disagreement with the statements concerning standardized test results and the "value-added" assessment for all age groups.

The respondents' years of experience were compared to their attitudes toward the "value-added" assessment as well. The same statements were used with the same scale from strongly disagree to strongly agree with a neutral position in the center. The results of that comparison are shown in Table 28. The same pattern emerged as was seen in the previous analysis. A large majority in all experience groups strongly disagreed with the statements, particularly those regarding the use of test results as a component of teacher evaluation, the validity and reliability of using test results to assess teacher performance, and the accuracy of a standardized test representing the breadth of students' learning. The greatest disagreement appeared with the statement concerning the reliability of tests to assess teacher performance with 65.2% strongly disagreeing. The statement concerning the validity of using students' test results for teacher assessment was next, with 64.8% strongly disagreeing.

These responses, more than any others, pointed out the attitudes of the educators surveyed toward standardized tests and the use of standardized test results as a component of teacher evaluation. As evidence, the educators surveyed were invited to respond to an open-ended question at the end of the survey questionnaire and comment on any issue raised in the instrument. Eighty-two comments were received, and of those, 72 were negative regarding standardized tests in general, the state-mandated T-CAP test, or, specifically, the use of standardized test scores as one component of teacher evaluation.

Table 28

Attitudes toward "Value-Added" Assessment by Years' Experience

	Percentage ^a of Responses					
Items ^b	1 ^c	2	3	4	5	N
1. Should be a component of teacher evaluation:						
< 5	45.0	30.0	25.0	0.0	0.0	20
5 - 10	77.1	11.4	8.6	2.9	0.0	35
11 - 15	81.8	9.1	9.1	0.0	0.0	33
16 - 20	67.7	19.4	9.7	0.0	3.2	31
21 - 25	47.6	28.6	21.4	0.0	2.4	42
26 - 30	44.4	33.3	7.4	11.1	3.7	27
> 30	42.9	14.3	19.0	9.5	14.3	21
Missing						13
Total	59.8	20.6	13.9	2.9	2.9	222
2. Are a valid means of assessing teacher performance:						
< 5	65.0	30.0	0.0	5.0	0.0	20
5 - 10	74.3	17.1	8.6	0.0	0.0	35
11 - 15	78.8	18.2	0.0	3.0	0.0	33
16 - 20	70.9	25.8	3.2	0.0	0.0	31
21 - 25	57.1	28.6	14.3	0.0	0.0	42
26 - 30	59.3	18.5	11.1	7.4	3.7	27
> 30	45.5	22.7	18.2	4.5	9.1	22
Missing						12
Total	65.3	22.9	8.1	2.4	1.4	222
3. Are a reliable means of assessing teacher performance:						
< 5	55.0	40.0	5.0	0.0	0.0	20
5 - 10	77.1	17.1	5.7	0.0	0.0	35
11 - 15	81.8	15.2	0.0	3.0	0.0	33
16 - 20	70.9	16.1	9.7	0.0	3.2	31
21 - 25	66.7	23.8	9.5	0.0	0.0	42
26 - 30	55.6	25.9	44.1	3.7	3.7	27
> 30	36.4	31.8	18.2	9.1	4.5	22
Missing						12
Total	65.7	22.9	8.1	1.9	1.4	222

Table 28 (continued)

	Percentage ^a of Responses					
Items ^b	1 ^c	2	3	4	5	N
4. Can indicate teacher effects on students' achievements:						
< 5	35.0	30.0	30.0	5.0	0.0	20
5 - 10	60.0	17.1	14.3	8.6	0.0	35
11 - 15	57.6	27.3	12.1	3.0	0.0	33
16 - 20	54.8	16.1	19.4	9.7	0.0	31
21 - 25	38.1	31.0	19.0	11.9	0.0	42
26 - 30	40.7	14.8	25.9	14.8	3.7	27
> 30	18.2	36.4	18.2	18.2	9.1	22
Missing						12
Total	45.3	24.3	19.0	10.0	1.4	222
5. Represent an accurate picture of the breadth of students' learning:						
< 5	45.0	30.0	20.0	5.0	0.0	20
5 - 10	68.6	17.1	14.3	0.0	0.0	35
11 - 15	60.6	27.3	9.1	3.0	0.0	33
16 - 20	64.5	22.6	9.7	3.2	0.0	31
21 - 25	57.1	21.4	16.7	4.8	0.0	42
26 - 30	44.4	29.6	18.5	0.0	7.4	27
> 30	22.7	36.4	22.7	13.6	4.5	22
Missing						12
Total	54.3	25.2	15.2	3.8	1.4	222

Table 28 (continued)

Items ^b	Percentage ^a of Responses					N
	1 ^c	2	3	4	5	
6. Reflect students' achievement gains:						
< 5	25.0	35.0	35.0	5.0	0.0	20
5 - 10	51.4	25.7	20.0	2.9	0.0	35
11 - 15	42.4	27.3	27.3	3.0	0.0	33
16 - 20	41.9	29.0	22.6	3.2	3.2	31
21 - 25	28.6	26.2	35.7	4.8	4.8	42
26 - 30	29.6	22.2	33.3	11.1	3.7	27
> 30	18.2	27.3	18.2	22.7	13.6	22
Missing						12
Total	35.2	27.1	27.6	6.7	3.3	222

^aPercentages may not add to 100% due to rounding

^bItems were statements beginning with "Standardized test results. . ."

^cResponses were from 1= Strongly Disagree to 5 = Strongly Agree

Following are some of the comments received:

The use of standardized test scores to evaluate teachers is unfair. Students should never be allowed to social[ly] pass grades. Each grade should have a standard of passing. There should be special school rooms and classes in each school for students who cannot pass. Have a class on test taking skills, test making, etc.

There are too many other factors involved in student learning to evaluate teachers according to standardized scores. Often tests do not evaluate learning.

Teachers should be held accountable for student's test scores when teachers carry those students *in utero* and when those students go home with those teachers every night.

Student's performances/abilities are not always accurately reflected by TCAP scores. All testing is not equal in all school districts. One district nearby had teachers who boasted of seeing tests the day before students were tested. Students were given answers to the test at that time. Why not use national tests? Because jobs have been created to develop the TCAP. Teacher uprising is on the way!

. . . Teacher evaluation based on standardized testing is bull manure. Present tests are not good enough, do not measure the right things, and are not in the proper format to measure a 'real' education. When (if????) we can make a 'proper' test; it's [sic] main purpose should be 'student' accountability not teacher accountability (and absolutely not teacher evaluation). We have tried educational reform based on everyone's accountability except the student. We test the student, then try to lay blame for poor results on everyone and everything except the student. The only educational reform that is going to have meaningful results is the one where student accountability is the 'main' part. . . .

Until testing is done professionally and all follow the same rules-test scores mean very little. I see too many high test scores, but students can't perform, therefore teacher evaluation based on test results will never work. TCAP's is [sic] a poor instrument and should be discontinued.

All comments to open-ended questions are reproduced in Appendix E.

To ascertain the effect the new evaluation requirements might have on teachers, respondents were asked to what degree they had changed or intended to change certain

aspects of their teaching in regard to the new requirements. Those respondents who were not classroom teachers omitted this question, leading to an N of less than 222. Those aspects of teaching behavior listed were classroom manner, course content, curriculum, methods of evaluation or assessment, and teaching methods; and respondents were asked to respond on a five point scale from "Never" to "Greatly." A response of "Don't Know" was also provided. The number of respondents and the means of the responses are shown in Table 29.

Teachers indicated that they would change or had changed the course content and curriculum somewhat. Also, they reported their method of evaluation and teaching methods have undergone some modifications. As the question posed, these changes would seem to be a result of the new teacher evaluation requirements. If that were so, it would seem that the means for the changes should be higher than seen. Perhaps the effects of the new requirements have not yet made themselves felt, and teachers have not yet shown a willingness to change to accommodate them. Or, teachers could feel very confident in what they do and feel that no change on their parts will be required. Asking the same question later after the evaluation process has been operative for some time could possibly yield different results.

Table 29

Degree of Change of Teaching Aspects

Items	Responses	
	N ^a	Mean ^b
Course content	206	2.55
Curriculum	206	2.48
Methods of evaluation/assessment	206	2.45
Teaching methods	206	2.36
Classroom manner	206	2.00

^aNumber responding (of the 222 participants)

^bResponses were from 1= Never to 5 = Greatly with 3 = Somewhat

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Summary

The problem under study was the need for teachers and administrators in the state of Tennessee to understand as much as possible of the science of tests and measurements and standardized tests as utilized by educational institutions and test manufacturers. The purpose of the study was to determine the perception of the general level of knowledge and attitudes that currently exist among teachers and administrators in the state of Tennessee about standardized testing, assessment, evaluation, and utilization of standardized test results. This study investigated the current perceptions of teachers and administrators concerning their knowledge of psychometrics and their attitudes toward standardized tests and the use of student standardized test scores as a component of teacher evaluation. The study attempted to answer the following questions:

1. How have student standardized test data been utilized by teachers and administrators in Tennessee?
2. What is the current level of perceived understanding about basic testing and measurement terms and processes by professional educators in Tennessee?
3. What training regarding tests and measurement have teachers and administrators received in their professional preparation?
4. What is the prevalent attitude towards using student standardized test data in teacher evaluations among practitioners in Tennessee?

The population for the study was composed of teachers and administrators in the state of Tennessee randomly selected from a list maintained by the Tennessee Education Association. To prevent any bias of the results, the list was composed of all current

teachers in the state, not only members of the association.

A survey instrument was mailed to the total sample along with a stamped, self-addressed return envelope. Because of assurances of confidentiality of replies, initially no record was kept of returned questionnaires. When the number of original returns was inadequate, a complete second mailing was made four months after the first mailing. Records of returned replies were kept, and two weeks after the second mailing, a postcard was mailed to encourage non-respondents to complete and return the survey questionnaire. Three weeks after the postcard was mailed, a third mailing was sent to those remaining non-respondents to elicit additional replies. Several weeks after the third mailing, telephone calls were made to those non-respondents whose telephone number could be found. Those contacted were asked demographic questions only to determine if they matched the profile of those who had responded to the survey questionnaires. The resultant data were analyzed using the Statistical Package for the Social Sciences (SPSS) program running on the VAX/VMS cluster at the University of Tennessee, Knoxville. The data were analyzed for frequency counts, descriptive statistics (means and number responding), and specific cross tabulations between those items that related to the research questions.

The following findings were made:

1. Student standardized test data have been most frequently utilized by teachers and administrators in the state of Tennessee for the purpose of diagnosing skill problems among students; a secondary use was for communicating to students' parents. The mean for each item was above a median score of 3.0 on a scale of 1.0 to 5.0, at 3.62 and 3.40 respectively. This use of standardized test scores for these purposes is in keeping with earlier studies, notably Dorr-Bremme and Herman in 1986 and Marso and Pigge in 1989 and 1992, but was not the most frequent purpose as identified by those studies. Other uses as identified by Tennessee teachers and administrators above the median score of 3.0 on a scale of 1.0 to 5.0 were those of assigning students to classes or programs,

curriculum planning, and informing the public. The majority of teachers with less than 5 years' experience and those with greater than 16 years' experience stated that standardized test data should never be used as a component of teacher evaluation. Given the predominately negative views expressed by many of the respondents to the open ended question, it was surprising that this use was not the lowest ranked. Allocation of funds was given the lowest ranking, indicating that, at least for the respondents, this was not a predominate use of standardized test scores.

When discussing standardized test scores with others, the respondents indicated that they discussed test score results most with other classroom teachers. This is not surprising in that teachers tend to talk among themselves more than not, and in a great deal of their conversation, they tend to discuss students and students' progress. In doing so, it seems natural that they would refer to test score data as a portion of that discussion. The respondents also indicated that test score results were discussed more frequently with parents than with the principal or the school's guidance department. That, too, is not surprising as teachers, in their discussions about students with parents, would relay test information to the parents as a means of defining the achievement of the student. That test scores are not discussed with the guidance office indicates a gap in the communication process between those that should have the students' interests in mind. Test score results were least discussed with those at the district office.

The respondents stated that the most-used basis for making educational decisions about students was their own observation and students' classwork with a mean of 4.61 on a scale from 1.0 to 5.0. The results of teacher-made tests was the next most frequent, followed by the results of other special placement tests. At the same time, teachers stated that they did not find standardized tests particularly valid or reliable for assessing student achievement; instead, as stated above, they preferred to use their own materials to make educational decisions concerning their students. Interestingly, while

the most common use of standardized tests as stated by educators was in the area of diagnosing student problems, when asked if they thought standardized tests were a reliable method for diagnosing those problems, a majority, 80.7%, said they were not. This suggests that many educators view standardized tests as virtually useless.

2. The level of perceived understanding of basic testing and measurement terms and processes by professional educators in Tennessee is slightly above an intermediate level. On a scale of from 1.0 to 5.0, the mean of the *perceived* knowledge of testing terms was 3.63 as identified by this study. The current level of use of basic testing and measurement terms and processes by professional educators in Tennessee as identified by this study is below an intermediate level at a mean of 2.32 on a scale of from 1.0 to 5.0. The above average ranking of respondents' perceptions of the supplied terms suggested a level of knowledge that, while not poor, could be improved. The respondents indicated that their perception of a term, how well they could relate or explain it to others was higher than their use of the term. The latter ranking, that of term usage, tends to be a stronger indicator of understanding in that those who have appropriated a term, i.e. a concept, would use it more frequently as part of their working vocabulary as opposed to others who have not. While it may be true that they simply may have no occasion to use the terms, true professionalism would indicate the opposite. When one appropriates a concept, they internalize it, and it then becomes a part of their everyday professional life. Simple knowledge or recognition of a term is not appropriation; it should be only part of a larger process.

Stiggins in 1991 stated that the status of assessment literacy in this country was poor and getting poorer. VanLeirsburg and Johns also decried the state of assessment literacy among future and practicing teachers in their 1991 study. Based on the data, the perception of knowledge of terms was higher compared to the usage of the same term yet was still only at an average level as identified by this study with the term usage at a lower level. This finding corresponds with that of Stiggins and VanLeirsburg. Not

unexpectedly, in all instances, the level of the perceived ability of the respondents to define a term was greater than the use of the term. In some cases, over a point and a half difference existed between the scores. The difference between the perception of a term and its use was found to be significant using a one-tailed t-test, $\alpha=.05$, $df=(\text{varies from } 422 \text{ to } 416)$, critical value=1.645 for all terms and their uses. The below intermediate ranking of term usage by educators in the study suggests that while the perceived knowledge is higher than what was considered average by this study, assessment literacy based on term usage was considered lower than what was considered average by this study.

As expected, as the respondents' age and years' experience increased, so did the perception of the terms associated with standardized testing and the science of tests and measurements. A comparison of the perception of term usage and the respondents' age and years' experience showed that as both age and years' experience increased, so did the perception of terms associated with testing and standardized tests. When an educator's age and years' experience increased, the perception of terms linked with testing increased, showing that both have an effect on an educator's perception of knowledge. That perception comes from the more frequent use of results from standardized tests and the experiences that teachers undergo.

The most frequent context of the use of testing terms as reported by the respondents in this study was in the special education M-Team--or Multi-Disciplinary Team--meeting. Here, respondents indicated that the M-Team meeting was the most prevalent context for the use of terms associated with standardized testing. Because students are normally assigned to special education because of scores on standardized tests, it was not unusual to find this as the most prevalent context. The next most-frequent context was reported as parent-teacher conferences. As stated earlier, when teachers and parents interact in meetings, one frequent assessment measurement that is quoted is the student's score on his/her most recent standardized test. The least-

mentioned context of term use was the conference between teacher and student. This perhaps suggests that a teacher does not feel that a student can interpret standardized test scores correctly, and therefore the information could possibly be either meaningless to the student or damaging to the student's self-concept.

3. The training regarding tests and measurements that teachers and administrators have received in their professional preparation has been fragmented and infrequently utilized. Formal training in psychometrics was split between the undergraduate and graduate levels and among five or more departments within a university/college. Indications of the use of the material studied in such classes was below an average ranking of 3.0 on a scale of 1.0 to 5.0 at 2.26 as identified by this study. Over 70% of the respondents stated that they had taken at least a partial course in tests and measurements, yet of those, only 10% stated they used material gained in those courses frequently. VanLeirsburg and Johns said in their 1991 study, "With the pervasive and increasing use of standardized tests, both preservice and inservice teachers will need to increase their assessment literacy in order to administer standardized tests correctly, prepare students to take tests, and make informed decisions about test scores" (p. 11). The low incidence of indication of use of learned material could suggest that that information might be deemed irrelevant by the respondents; that is, that the material learned in such courses could not be correlated with the respondent's professional career. Educators, needing to see a connection between information gained in college courses or that gained by individual study and their work, might not establish a nexus between that material presented by a college or university and their position. This could explain the low incidence of use of supposedly-learned course material.

When asked about learning about tests and measurements on their own initiative, only 37.4% stated they had used other methods to learn about the topic. Again, if educators do not recognize a connection between learned material and their job, it is

unlikely that any transfer will take place. Concurrently, if no reinforcement is offered between the learning of material and the application, the learning becomes unimportant. As far as individuals were concerned, the most frequently-used method of gaining information was a personal study of books, journals, or articles related to testing and assessment. Respondents were not asked how frequently they used the material learned in this fashion, the assumption being that if they took the time and effort to involve themselves, some learning and application would take place. Conversely, respondents were asked how frequently their school or district offered in-service or additional training about testing topics. Respondents indicated that their school or district only sometimes discussed methods to improve test scores or methods to strengthen instruction in the specific areas where test scores showed weaknesses with a mean score of 3.0 on a scale of 1.0 to 5.0. The lowest ranking for this item was one that dealt with the assisting of teachers in specific testing and evaluation techniques, with the mean score being 2.33, a below average ranking of 3.0 on a scale of from 1.0 to 5.0. This would tend to suggest that districts do not place a high emphasis on assisting their teachers with that information that would increase their knowledge about testing procedures and standardized testing.

Respondents were also asked how frequently they utilized the technical manual that accompanies--or should accompany--each standardized test. Those who have been exposed to testing procedures and processes would recognize that a technical manual accompanies each standardized test. Contained within this manual are the statistics that define the test including the norming curves. To deny the existence of the technical manual for a standardized test indicated a serious lack of knowledge of testing procedures. Almost two-thirds of the respondents stated they never or almost never used the manual, and one did not even know it existed.

4. The prevalent attitude towards using student standardized test data in teacher evaluations among practitioners in Tennessee was overwhelmingly negative. One major

aspect of this study dealt with the attitudes that educators held toward standardized testing in general and specifically toward using standardized test scores as a component of teacher evaluations. As stated above, because the state of Tennessee has chosen to use students' standardized test scores as a component of teacher evaluation in its "value-added" assessment program, it was important to address the attitudes of educators towards the medium by which they were to be evaluated. Standardized testing has been a part of the educational milieu for many years. Its uses have been many and varied. Now with the advent of using test data as a part of teacher evaluations, it became even more important to investigate those attitudes.

The respondents indicated that they agreed that standardized tests did their best when they were used to to test the cognitive domain of student learning. On a scale of 1.0 to 5.0, the mean ranking of that item was 3.02, not an overwhelming affirmation of belief, but one that was slightly above average. Slightly below that at a mean of 2.99 was the belief that standardized tests test lower-order thinking skills. As stated above, teachers did not feel that standardized tests were either a reliable or valid means of assessing student achievement with mean scores of 2.42 and 2.40 respectively. This suggested that educators do not place a large amount of faith in the ability of standardized tests to display the breadth of student learning. The lowest ranking was afforded the statement that standardized tests tested the affective domain, at 2.22. Teachers did not feel that standardized tests could adequately measure students' emotions as compared to their intellect.

The respondents tended to disagree more with statements that dealt with the use of standardized tests for different uses. The highest ranking was earned by the statement that standardized test results reflected students' achievement gains with a mean score of 2.16 on a scale of 1.0 to 5.0. The lowest rankings were associated with the statements that standardized tests were a valid and reliable means of assessing teacher performance at a mean score of 1.52 and 1.50 respectively. Educators were were firmly against the

use of standardized tests as a component of teacher evaluations, ranking that use at a mean of 1.68 on a scale of from 1.0 to 5.0. Additionally, over 90% percent of the open-ended comments about any one topic on the survey instrument were those dealing with the use of standardized tests for teacher evaluation. The majority were against that specific use of standardized test data. One teacher stated, "There are too many other factors involved in student learning to evaluate teachers according to standardized scores. Often tests do not evaluate learning." Other comments were in the same vein.

Conclusions and Discussion

1. This study has identified the most frequent use of student standardized test data in the state of Tennessee as diagnosing student problems. When student test data are used primarily in this manner, it limits the effectiveness of the utilization of test data to a narrow field. Test data, according to the respondents, are not being utilized for its primary function as others have identified. While distinguishing student problem areas is an important function, it is not the best use for the data. Herman and Dorr-Bremme (1984) found that student concerns rated near the middle of their findings for the uses of standardized test data. Darling and Hammond (1991) identified several uses for test scores as they applied to students, such as tracking, retention, rewards, or sanctions, but these uses were not the most frequent. As George Madaus stated in 1991, it is true that using student standardized test scores has long been used to sort and classify students. To use standardized test scores most frequently for this purpose, especially when teachers stated that they did not rely on test scores as the best indicator of students' performance, is of dubious benefit to both the student and teacher alike.

Student test data certainly have many uses. Perhaps the best overall function for student standardized test scores, as others have suggested, is in the arena of planning. When student test scores are used in a limited fashion, as indicated by this study, the

scope of the usefulness of the data is also limited. Successful curriculum planning requires input from many sources, student test data being just one of those sources. Bracey (1991, 1992) and others (Linden & Linden, 1968; Madaus, 1991; Norman, 1980; Olejnik, 1979; Shepard, 1989; Wiggins, 1991) have stated that the multiple-choice, forced-answer form of tests currently administered does not address all areas of student performance. As Bracey (1992) said, "All tests measure only what students know and can do now; . . ." (p. 414). As such, educators who use standardized tests merely for detecting student skill problems, especially in higher-order thinking areas, are not using tests for their best purpose. As such, test scores can point out weaknesses, problem areas for the student and possibly assist in planning for strategies to assist that student. But if this becomes the sole purpose for test score usage, it is not the most advantageous use.

In this era of an increasing call for accountability for students' achievement, educators must use all the data available to them to make informed decisions. Teachers should be aware of the information available to them from students' test scores and use that information appropriately. If a teacher examines a student's testing record and finds it can only point out problems in that student's performance, then much is lost. Yet to be able to correctly interpret that record, the teacher, guidance counselor, principal, and system-wide administrator must be able to understand what the record is saying. Only with that knowledge, the knowledge of what the results mean, can better uses be found for the data.

2. The gap between educators' perceptions of testing terms and the use of those terms in professional situations indicates a lack of test awareness. Because this study attempted to determine the perception of knowledge rather than actual level of knowledge, no firm conclusion can be drawn concerning the actual level of knowledge of educators in Tennessee. Yet the low ranking of the usage of testing terms in professional situations indicates a lack of test awareness among educators. In instituting the method

of value-added assessment for its teachers, the state of Tennessee has implicitly made the assumption that its practitioners understand the practice and application of those aspects concerned with testing and standardized tests. While that is not necessarily an invalid assumption, this study indicates that the level of knowledge of current practitioners is based on the use of testing terms in professional situations. Educators indicate they know what a term means, but do not use the terms in the frequencies that would display more than a superficial level of knowledge. The gap between the perception of knowledge and actual use may be due to the infrequency with which educators are called upon to display their knowledge about testing terms and procedures, or it may display an actual gulf between knowledge and practice. A person trained in a profession, be it medicine, law, or education, should be able to utilize the terminology of that profession in discussions of occupational concern with others; else how could they make themselves understood? If a person felt that he or she could define a term successfully, then that term would become part of his or her professional, working vocabulary. Yet, the indication of usage of all the terms was consistently lower than the indication of the respondent's ability to define it, leading to the finding that while the respondents stated they could define the term, it was not used in situations where the use of the term would be correct in context nor apparently had it been assimilated into one's working lexicon.

3. Of the methods offered to gain information about tests and measurements, the material learned in college courses, that learned by educators on their own initiative, and that offered by schools and districts are not being utilized to their best advantage.

The respondents in this study did not display a high degree of assessment literacy as a result of the course work in which they said they had been involved. This suggests that they perceive no relevancy between the course material and their professional careers. The large number of departments offering courses in tests and measurements at both the undergraduate and graduate levels might suggest wide variations in information

provided, focus of instruction, emphasis of material, and method of presentation.

Of the information offered to educators by their schools/districts, from the findings, it appears that the schools/districts have not placed a high priority on assisting their staff with information related to testing and assessment methods. This is another indication of the lack of utilization of information about standardized testing. Schools/districts, through their information gathering and dissemination systems, should be able to provide their staff members with current data concerning tests and measurements, assessment methods, and test score usage. Also, they should be able to provide in-service sessions that will enable their staff members to understand better the new "value-added" evaluation.

No matter by which of the various methods educators gained their information, through college courses or independent study, the findings indicated that it was not being utilized to help them in their profession. The low rankings given to term perception, term use, and utilization of learned course material suggest that no matter what the means of accumulation of information, utilization of that information was low. This indication of how well information is being learned and utilized indicates that knowledge about tests and measurements is, as previously stated, not being utilized to its best advantage.

4. Educator's attitudes toward standardized tests are doubtful and contradictory.

As indicated by the findings of this study, many educators are unsure about the effectiveness of standardized tests to measure students' achievements. The respondents indicated from strong disagreement to strong agreement on all items dealing with attitudes toward standardized tests and the use of test scores as a component of teacher evaluation. This suggests that a wide variety of conceptions and perceptions existed about both the science of tests and measurement and the purposes underlying testing programs. One reason for this uncertainty and confusion may lie in the amount of knowledge possessed by the respondent. Another reason for the doubtfulness displayed

could be misunderstanding of the testing process. As stated above, the respondents indicated a wide range of scores in their answers to the questionnaire items. In expressing these feelings, educators tended to contradict themselves as they indicated their responses. This also suggests that a strong lack of knowledge exists among practitioners in the state.

5. Educators are uneasy about the new evaluation requirements utilizing student standardized test scores. In this era of increasing accountability, administrators look for objective data to apply to performance evaluations of their personnel. Cook (1994) suggested that "When there are significant differences in teachers' effectiveness, it is feasible that an objective component of teacher evaluation based upon achievement test data can be developed" (p. 164). The state of Tennessee in its "value-added" assessment program has chosen to use students' standardized test data as a component of teacher evaluation, basing its decision on the work of Maclean and Sanders (1984, 1991). When asked about the use of student standardized test data for this purpose, a majority of the respondents replied that students' test scores should not be a part of teacher evaluations. The primary reason for this was the perception that students' performance on standardized tests was out of the teachers' control. Teachers' comments in this area were strongly negative.

Recommendations

The following are the recommendations from this study:

Recommendation 1: Further research should be conducted as to the specific knowledge of test awareness rather than the perception level of educators in Tennessee. Specific knowledge levels could support colleges of teacher preparation in their curricula plans to assist educators gain the knowledge they need.

Recommendation 2: Colleges of teacher preparation in the state of Tennessee

must incorporate additional training about tests and measurement in their teacher education and administrator training programs. Additionally, information relating specifically to the "value-added" assessment procedures must be included in all teacher and administrator preparation programs.

Recommendation 3: The state should institute further research into the effects of the "value-added" assessment procedures on both students and educators. Given the predominately negative attitudes of educators toward the new assessment requirements, more investigation into the results of the "value-added" assessment is definitely warranted.

Recommendation 4: More reliance on authentic assessment methods should be utilized in evaluating students rather than relying strongly on standardized test data. Because of the limitations of standardized testing, more and better evaluation methods should be utilized to correctly assess students and ensure their placement in the proper educational settings.

Recommendation 5: Those who administer the tests in the state must receive specific training in the tests they administer. While we do not expect our teachers to be experts at tests and measurement, we should expect them to know something about the tests they are giving. More information about standardized testing procedures, and in Tennessee, specific information about the T-CAP would be more than beneficial, it should be mandatory.

Recommendation 6: Schools and districts should increase their level of involvement of teachers in activities to help them extend their level of understanding about tests and measurement. For educators who are not involved in advanced course work through a college or university, the school district should be the major source of information. In addition to assisting educators to amplify their knowledge base, a district should supply information in the form of books and journals to assist its teachers in gaining the knowledge they need.

BIBLIOGRAPHY

Bibliography

- Airasian, P. W., & Madaus, G. (1974). Criterion-Referenced testing in the classroom. In R. W. Tyler & T. M. Wolf (Eds.), *Crucial Issues in Testing* (pp. 70-104). Berkeley, CA: McCutchan.
- Archbald, D. A., & Newmann, F. M. (1988). *Beyond standardized testing: Assessing authentic academic achievement in the secondary school*. Reston, VA: National Association of Secondary School Principals.
- Ashworth, K. H. (1989-1990, Winter). Standardized testing: A defense. *The College Board Review*, 10-13, 23-24.
- Beatty, W. H. (1969). *Improving educational assessment and an inventory of measures of affective behavior*. Washington, DC: Association for Supervision and Curriculum Development.
- Bowers, B. C. (1989). *Alternatives to Standardized Educational Assessment* (ERIC Digest Series Number EA 40). Eugene, OR: ERIC Clearinghouse on Educational Management. (ERIC Document Reproduction Service No. ED 312 773)
- Bracey, G. W. (1992, January). Getting smart(er) in school. *Phi Delta Kappan*, 73, 414-416.
- Bracey, G. W. (1991, October). Why can't they be like we were? *Phi Delta Kappan*, 73, 104-117.
- Brandt, R. (1989, April). On misuse of testing: A conversation with George Madaus. *Educational Leadership*, 46(7), 26-29.
- Brown, D. F. (1991, Fall). Are state tests effective measures of student abilities? Perspectives of teachers and principals. *Tennessee Education*, 21, 10-14.
- Brown, R. (1989). Testing and Thoughtfulness. *Educational Leadership*, 46(7), 31-33.
- Cohen, M. (1986). Designing state assessment systems. Commissioned paper cited in Appendix B of *The Nation's Report Card*. Washington, DC: Study Group on the National Assessment of Student Achievement. (ERIC Document Reproduction Service No. ED 279 670)
- Cook, D. L. (1984). *An objective component for the evaluation of teaching*. Unpublished doctoral dissertation, The University of Tennessee, Knoxville.
- Cornett, L. M., & Gaines, G. F. (1991). Linking performance to rewards for teachers, principals, and schools. *The 1990 SREB Career Ladder Clearinghouse Report*. Atlanta, GA: Southern Regional Education Board.

- Darling-Hammond, L. (1991, November). The implications of testing policy for quality and equality. *Phi Delta Kappan*, 73, 220-225.
- Dorr-Bremme, D., & Herman, J. L. (1986). *Assessing student achievement: A profile of classroom practices*. Los Angeles, CA: Center for the Study of Evaluation, University of California.
- Dyer, H. S., & Rosenthal, E. (1974). Assessing education at the state level. In R. W. Tyler & T. M. Wolf (Eds.), *Crucial Issues in Testing* (pp. 105-128). Berkeley, CA: McCutchan.
- Education Improvement Act of 1992*, Tennessee SB 1231 by Womack and HB 752 by Purcell. Nashville, TN: State of Tennessee.
- Elam, S. M., Rose, L. C., & Gallup, A. M. (1992, September). The 24th annual Gallup/Phi Delta Kappan poll of the public's attitudes toward the public schools. *Phi Delta Kappan*, 74, 41-53.
- Ewell, P. T., & Boyer, C. M. (1988, July-August). Acting out state-mandated assessment. *Change*, 20, 40-48.
- Goertz, M. E. (1988). *State educational standards in the 50 states: An update*. Princeton, NJ: Educational Testing Service.
- Haney, W., & Madaus, G. (1986). Effects of standardized testing and the future of the national assessment of educational progress. Working paper for the NAEP study group. Commissioned paper cited in Appendix B of *The Nation's Report Card*. Washington, DC: Study Group on the National Assessment of Student Achievement. (ERIC Document Reproduction Service No. ED 279 680)
- Herman, J., & Dorr-Bremme, D. W. (1984). *Testing and assessment in American public schools: Current practices and directions for improvement*. Los Angeles, CA: University of California, Center for the Study of Evaluation. (ERIC Document Reproduction Service No. ED 251 508)
- Jaeger, R. M. (1991, November). Legislative perspectives on statewide testing: Goals, hopes, and desires. *Phi Delta Kappan*, 73, 239-242.
- Johns, J. L., & Davis, S. J. (1991). *Perceptions of preservice and inservice teachers regarding test-taking procedures and test-wiseness programs*. (Literacy research report no. 4). Dekalb: Northern Illinois University, Curriculum and Instruction Reading Clinic.
- Koretz, D. (1988, Summer). Arriving in Lake Wobegon: Are standardized tests exaggerating achievement and distorting instruction? *American Educator*, 12(2) pp. 8-15, 46-52.
- Lewis, A. E. (1986). The politics of testing/assessment (or a chameleon in the

- classroom). Commissioned paper cited in Appendix B of *The Nation's Report Card*. Washington, DC: Study Group on the National Assessment of Student Achievement. (ERIC Document Reproduction Service No. ED 279 686)
- Lieberman, A. (1991, November). Accountability as a reform strategy. *Phi Delta Kappan*, 73, 219-220.
- Linden, K. W., & Linden, J. D. (1968). *Modern mental measurement: A historical perspective*. Boston, MA: Houghton Mifflin.
- Madaus, G. F. (1991, November). The effects of important tests on students: Implications for a national examination system. *Phi Delta Kappan*, 73, 226-231.
- Maeroff, G. I. (1992, January). Focusing on urban education in Britain. *Phi Delta Kappan*, 73, 352-358.
- Marso, R. N., & Pigge, F. L. (1989). *Staff development implications from a state-wide assessment of classroom teachers' testing skills and practices*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 312 309)
- Marso, R. N., & Pigge, F. L. (1992). *A summary of published research: Classroom teachers' knowledge and skills related to the development and use of teacher-made tests*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED 346 148)
- McLaughlin, M. W. (1991, November). Test-based accountability as a reform strategy. *Phi Delta Kappan*, 73, 248-251.
- McLean, R. A., Saunders, W. L., & Stroup, W. W. (1991). A unified approach to mixed linear models. *The American Statistician*, 45, 54-64.
- McLean, R. A., & Saunders, W. L. (1984). *Objective component of teacher evaluation--A feasibility study. Working Paper #199*. Knoxville, TN: College of Business Administration, University of Tennessee.
- Mehrens, W. A., & Lehmann, I. J. (1987). *Using standardized tests in education* (4th ed.). New York: Longman.
- National School Boards Association. (1977). *Standardized achievement testing*. Washington: Author.
- Norman, C. A. (1980). *Measurement and testing: An NEA perspective*. Washington: National Education Association.
- Olejnuk, S. F. (1979). *Standardized achievement testing programs viewed from the*

perspective of a non-measurement specialist. Paper presented at the annual meeting of the National Council of Measurement in Education, San Francisco, CA. (ERIC Document Reproduction Service No. ED 177 209)

- Roeder, H. H. (1973, Summer). Teacher education curricula--Your final grade is F. *Journal of Educational Measurement*, 10(2), 141-143.
- Ruddell, R. B. (1985, February). Knowledge and attitudes toward testing: Field educators and legislators. *The Reading Teacher*, 38(6), 538-543.
- Shepard, L. A. (1989, April). Why we need better assessments. *Educational Leadership*, 46(7), 4-9.
- South Carolina State Department of Education, Division of Public Accountability. (1989, February). *South Carolina Education Improvement Act of 1984, as Amended, 1985-1988*. Columbia, SC: Author.
- Stake, R. E. (1991, November). The teacher, standardized testing, and prospects of revolution. *Phi Delta Kappan*, 73, 243-247.
- Stiggins, R. J. (1991, March). Assessment literacy. *Phi Delta Kappan*, 72, 534-539.
- Tyler, R. W., & Wolf, T. M. (Eds.). (1974). *Crucial issues in testing*. Berkeley, CA: McCutchan.
- VanLeirsburg, P., & Johns, J. L. (1991). *Assessment literacy: Perceptions of preservice and inservice teachers regarding ethical considerations of standardized testing procedures* (Literacy research report no. 12). Dekalb: Northern Illinois University, Curriculum and Instruction Reading Clinic. (ERIC Document Reproduction Service No. ED 341 666)
- Wiggins, G. (1991, February). Standards, not standardization: Evoking quality student work. *Educational Leadership*, 48(5), 18-26.
- Wiggins, G. (1989, April). Teaching to the (authentic) test. *Educational Leadership*, 46(7), 41-47.

Secondary References

- Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.
- Benveniste, G. (1985, Fall). The design of school accountability systems. *Educational Evaluation and Policy Analysis*, 7(3), 261-280.
- Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1981). *Evaluation to improve learning*. New York, NY: McGraw-Hill.
- Boyer, E. (1988, January). School reform: Completing the course. *National Association of Secondary School Principals Bulletin*, 72, 61-68.
- Bracey, G. W. (1990, January). Teachers, thinking, and testing. *Phi Delta Kappan*, 71, 404-408.
- Bracey, G. W. (1987, May). Measurement-driven instruction: Catchy phrase, dangerous practice. *Phi Delta Kappan*, 68, 683-686.
- Brandt, R. M. (1987, April). *Improving the quality of student outcome data used in career ladder promotion decisions*. Paper presented at the annual meeting of the American Research Association, Washington, DC. (ERIC Document Reproduction Service No. ED 301 928)
- Cibulka, J. G. (1989, March). *State accountability reforms: Performance information and political power*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Cibulka, J. G. (1989, August). State performance incentives for restructuring: Can they work? *Education and Urban Society*, 21(4), 417-435.
- Cizek, G. J. (1991, May). Innovation or enervation? Performance assessment in perspective. *Phi Delta Kappan*, 72, 695-699.
- Cizek, G. J. (1991, October). Confusion effusion: A rejoinder to Wiggins. *Phi Delta Kappan*, 73, 150-153.
- Dillon, R. F., & Pellegrino, J. W. (1989). *Testing: Theoretical and applied perspectives*. New York, NY: Praeger.
- Ebel, R. L. (1977). *The uses of standardized testing*. Fastback Series # 93. Bloomington, IN: The Phi Delta Kappa Educational Foundation.
- Ferguson, G. A., & Takane, Y. (1989). *Statistical analysis in psychology and education* (6th ed.). New York: McGraw-Hill.
- Feuer, M. J., Fulton, K., & Morison, P. (1993, March). Better tests and testing

- practices: Options for Policy Makers. *Phi Delta Kappan*, 74, 530-533.
- Fink, A., & Kosecoff, J. (1985). *How to conduct surveys: A step-by-step guide*. Beverly Hills: SAGE Publications.
- Gay, R. L. (1981). *Educational research: Competencies for analysis and application* (2nd ed.). Columbus, OH: Charles E. Merrill.
- Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991, June-July). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20, 2-7.
- Hamm, R. W. (1988, Spring). Educational Evaluation: Theory and a working model. *Education*, 108, 404-409.
- Hansen, J. M. (1989, December). Outcome-based education: A smarter way to assess student learning. *The Clearing House*, 63, 172-175.
- Hawes, G. R. (1964). *Educational testing for the millions: What tests really mean for your child*. New York: McGraw-Hill.
- Herman, J. L., & Golan, S. (1990). *Effects of standardized testing on teachers and learning: Another look*. CSE Technical Report 334. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.
- Huff, K., and others. (1989, March). *A comparative analysis of achievement gains and related school characteristics*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED 314 423)
- Kellaghan, T., Madaus, G. F., & Airasian, P. W. (1982). *The effects of standardized testing*. Boston, MA: Kluwer-Nijhoff Pub.
- Killiffer, J. (1989). *Effective schools correlates and achievement gain: A relational study*. Unpublished doctoral dissertation, The University of Tennessee, Knoxville.
- McAlister, D. (1991). Education legislation. *Tennessee School Boards Journal*, 8,(4), 6-9.
- Mandeville, G. K., & Anderson, L. W. (1986, April). *A study of the stability of school effectiveness measures across grades and subject areas*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA. (ERIC Document Reproduction Service No. ED 271 497)
- Mazzarella, J. A. (1975). *Accountability and testing*. Arlington, VA: National Association of Elementary School Principals.

- National Education Association. (1977). *Standardized testing issues: Teachers' perspectives*. Washington, DC: Author.
- Porter, A. C. (1983). The role of testing in effective schools. *American Education*, 25-28.
- Renfrow, D. D., & Kromrey, Jeffrey D. (1990). *Descriptions and uses of content-specific pedagogical items*. Paper presented at the annual meeting of the National Council on Measurement in Education, Boston, MA. (ERIC Document Reproduction Service No. ED 322 175)
- Rudman, H. C. (1985, February). *Testing beyond minimums. Occasional paper #5*. Paper presented at the Michigan School Testing Conference, Ann Arbor, MI. (ERIC Document Reproduction Service No. ED 268 182)
- Ryan, J. M., & Rowzie, G. G. (1987, May). *A manual for the student achievement component of the teacher incentive program*. Columbia, SC: College of Education, University of South Carolina.
- Shepard, L. A. (1991, November). Will national tests improve student learning? *Phi Delta Kappan*, 73, 232-238.
- Study Group on the National Assessment of Student Achievement. (1986). *The Nation's Report Card*. Washington, DC: Author.
- Teacher Evaluation Programs in SREB States*. (1990). Atlanta, GA: Southern Regional Education Board. (ERIC Document Reproduction Service No. ED 334 655)
- Travers, R. M. W. (1981). Criteria of good teaching. In J. Millman (Ed.), *Handbook of teacher evaluation* (pp. 14-22). Beverly Hills: SAGE Publications in cooperation with the National Council on Measurement in Education.
- Wiggins, G. (1991, May). A response to Cizek. *Phi Delta Kappan*, 72, 700-703.
- Worthern, B. R., & Spandel, V. (1991, February). Putting the standardized test debate in perspective. *Educational Leadership*, 48(5), 65-69.

APPENDICES

Appendix A

Survey Questionnaire for Tennessee Educators about Standardized Test Knowledge and Usage

1. Your age: 1 _____ 20 - 25
 2 _____ 26 - 30
 3 _____ 31 - 35
 4 _____ 36 - 40
 5 _____ 41 - 45
 6 _____ 46 - 50
 7 _____ > 50
2. Your years experience in education: 1 _____ < 5 years
 2 _____ 5 - 10 years
 3 _____ 11 - 15 years
 4 _____ 16 - 20 years
 5 _____ 21 - 25 years
 6 _____ 26 - 30 years
 7 _____ > 30 years
3. Grades taught currently: 1 _____ Early Elementary (K-3)
 2 _____ Elementary (K-6)
 (Check one) 3 _____ Middle School (6-8)
 4 _____ Junior High (7-9)
 5 _____ High School (9-12)
 6 _____ Senior High (10-12)
 7 _____ System-wide
4. Highest educational level: 1 _____ BS/BA
 (Check one) 2 _____ MS/MA
 3 _____ MS/MA+45
 4 _____ Ed.S
 5 _____ Ph.D/Ed.D
 6 _____ Other
5. Currently working on an advanced degree (Check one): 1 _____ YES
 2 _____ NO
 3 _____ UNDECIDED
6. Current assignment: 1 _____ Classroom teacher
 (Check one) 2 _____ Principal
 3 _____ System-wide administrator
 4 _____ Other: _____

7. Based on your current assignment, please indicate how often standardized test scores are used for the following purposes. For a classroom teacher, consider your own classroom; for a principal, consider your school; for an administrator, consider your school district. Circle the most appropriate number (DK = Don't Know):

	Never	Sometimes	Frequently	
1 Allocation of funds	1	2	3	4 5 DK
2 Assigning students to classes/programs	1	2	3	4 5 DK
3 Communicating to parents	1	2	3	4 5 DK
4 Curriculum planning	1	2	3	4 5 DK
5 Diagnosing skill problems	1	2	3	4 5 DK
6 Informing the public	1	2	3	4 5 DK
7 Policy formation	1	2	3	4 5 DK
8 Reporting to district	1	2	3	4 5 DK
9 Student promotion/retention	1	2	3	4 5 DK
10 Teacher evaluation	1	2	3	4 5 DK

8. How frequently do you discuss standardized test results with the following? Circle the most appropriate number (DK = Don't Know):

	Never	Sometimes	Frequently	
1 District Office	1	2	3	4 5 DK
2 Guidance department	1	2	3	4 5 DK
3 Other Classroom teachers	1	2	3	4 5 DK
4 Parents	1	2	3	4 5 DK
5 Principal	1	2	3	4 5 DK
6 Students	1	2	3	4 5 DK

9. To what degree do you prefer to use the following to make educational decisions concerning students? Circle the most appropriate number:

	Never	Sometimes	Frequently	
1 Curriculum materials' placement test results	1	2	3	4 5
2 Own observations and students' classwork	1	2	3	4 5
3 Previous teachers' comments, reports, grades	1	2	3	4 5
4 Results of other special placement tests	1	2	3	4 5
5 Results of teacher-made tests	1	2	3	4 5
6 Students' other standardized test results	1	2	3	4 5
7 Students' scores of T-CAP tests	1	2	3	4 5

** Please use the following scale to respond to all of the items in the next two questions:

Strongly Disagree					Strongly Agree
1	2	3	4	5	

10. To what degree do you agree or disagree with the following statements concerning standardized tests?

1. _____ a) They provide a reliable means of assessing student achievement.
2. _____ b) They provide a valid means of assessing student achievement.
3. _____ c) They provide a reliable method for diagnosing student problems.
4. _____ d) They test the cognitive domain.
5. _____ e) They test the affective domain.
6. _____ f) They test higher-order thinking skills.
7. _____ g) They test lower-order thinking skills.

11. According to the BEP, student test results converted to a 'value-added assessment' will be used as one component of evaluation for teachers beginning April 1, 1995. To what degree do you agree or disagree with the following statements concerning the use of 'value-added' test results as a component of teacher evaluation?

1. _____ a) Standardized test results should be a component of teacher evaluations.
2. _____ b) Standardized test results are a valid means of assessing teacher performance.
3. _____ c) Standardized test results are a reliable means of assessing teacher performance.
4. _____ d) Standardized test results can indicate teacher effects on students' achievements.
5. _____ e) Standardized test results represent an accurate picture of the breadth of students' learning.
6. _____ f) Standardized test results reflect student's achievement gains.

12. To what degree do you intend to change, or have you changed, the following aspects of your teaching with regard to the new evaluation requirements? Circle the most appropriate number (DK = Don't Know):

	Never	Somewhat			Greatly		
1 Classroom manner	1	2	3	4	5		DK
2 Course content	1	2	3	4	5		DK
3 Curriculum	1	2	3	4	5		DK
4 Methods of evaluation/assessment	1	2	3	4	5		DK
5 Teaching methods	1	2	3	4	5		DK

13. If you were asked to define each term listed below, how well could you explain it to someone who was unfamiliar with testing and measurement? Circle the most appropriate number (DK = Don't Know):

	Not Well	Fairly Well			Very Well		
1 Adjusted Score	1	2	3	4	5		DK
2 Criterion-referenced test	1	2	3	4	5		DK
3 Grade equivalents (GE)	1	2	3	4	5		DK
4 Mean	1	2	3	4	5		DK
5 Median	1	2	3	4	5		DK
6 Mode	1	2	3	4	5		DK
7 Normal distribution	1	2	3	4	5		DK
8 Norm-referenced test	1	2	3	4	5		DK
9 Percentile	1	2	3	4	5		DK
10 Percentile rank	1	2	3	4	5		DK
11 Raw score	1	2	3	4	5		DK
12 Reliability	1	2	3	4	5		DK
13 Scale Score	1	2	3	4	5		DK
14 Standard Score	1	2	3	4	5		DK
15 Stanine	1	2	3	4	5		DK
16 Statistical process control	1	2	3	4	5		DK
17 Validity	1	2	3	4	5		DK

14. For each term given below, rate your level of use, i.e. how often do you use each term in your conversations with other teachers, administrators, parents, students, etc. Circle the most appropriate number:

	Never Use		Sometimes Use		Frequently Use
1 Adjusted Score	1	2	3	4	5
2 Criterion-referenced test	1	2	3	4	5
3 Grade equivalents (GE)	1	2	3	4	5
4 Mean	1	2	3	4	5
5 Median	1	2	3	4	5
6 Mode	1	2	3	4	5
7 Normal distribution	1	2	3	4	5
8 Norm-referenced test	1	2	3	4	5
9 Percentile	1	2	3	4	5
10 Percentile Rank	1	2	3	4	5
11 Raw score	1	2	3	4	5
12 Reliability	1	2	3	4	5
13 Scale Score	1	2	3	4	5
14 Standard Score	1	2	3	4	5
15 Stanine	1	2	3	4	5
16 Statistical process control	1	2	3	4	5
17 Validity	1	2	3	4	5

15. What is (are) the context(s) when you use these terms? Check all that apply:

- 1 _____ a) Central office meetings
 2 _____ b) Faculty meetings
 3 _____ c) Multi-disciplinary (M-Team) meetings
 4 _____ d) Parent-teacher conferences
 5 _____ e) Student conferences
 6 _____ f) Teacher evaluation conferences
 7 _____ g) Workshops/Seminars
 8 _____ h) Other _____?

16. Have you ever taken a specific course in tests and measurements?

- 1 _____ YES, a complete course 2 _____ YES, a partial course (some aspects of testing)
 3 _____ NO 4 _____ DONT REMEMBER

**** If the answer is either YES answer, please complete the following; if the answer is NO or DONT REMEMBER, please skip to question #19.**

17. Use the following responses to describe the course(s) in tests and measurements you have taken; please check all that apply:

- a) Level: 1 _____ Undergraduate
 2 _____ Graduate
 b) Department: 3 _____ Business
 4 _____ Curriculum and Instruction
 5 _____ Educational Administration and Supervision
 6 _____ Educational Psychology
 7 _____ Liberal Arts
 8 _____ Other: _____?

18. How frequently do you utilize material from that (those) course(s): (Circle the most appropriate number)?

Not very Frequently		Frequently		Very Frequently
1	2	3	4	5

19. Have you ever undertaken work independently to increase your level of understanding about tests and measurements?

1_____ YES 2_____ NO 3_____ DONT REMEMBER

20. If the answer to #19 is YES, what method did you employ (books, journal articles, other instruction, etc.; please list all you know)?

21. How frequently have you utilized the technical manual that accompanies each standardized test?

Never					Often
	1	2	3	4	5

22. How often during the year does your school/district engage in the following activities related to testing with teachers?

	Never				Often
1 Holds staff meetings to review test scores	1	2	3	4	5
2 Discusses methods to improve test scores	1	2	3	4	5
3 Discusses methods to strengthen instruction in the specific areas where test scores show weakness	1	2	3	4	5
4 Assists teachers in specific testing and evaluation techniques	1	2	3	4	5
5 Holds in-service specifically devoted to tests and measurements	1	2	3	4	5

23. Have you ever taught in a state or district where student standardized test results were used as a component of teacher evaluation?

1_____ YES (If YES, Where? _____)
2_____ NO 3_____ DONT REMEMBER

24. If the answer to #23 is YES, please complete the items in the following two questions; if the answer is NO or DONT REMEMBER, please skip to #26:

Did you think that the use of students' standardized test results as a component of teacher evaluation was (Circle the most appropriate number; UN = Undecided):

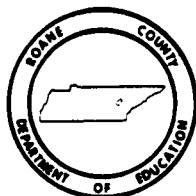
	Not at all		Somewhat		Very	
1 Accurate	1	2	3	4	5	UN
2 Consistent	1	2	3	4	5	UN
3 Fair	1	2	3	4	5	UN
4 Unbiased	1	2	3	4	5	UN
5 Valid	1	2	3	4	5	UN

25. If you were to teach in that state or district again, would you want to be evaluated using students' standardized test results as a component of your evaluation?

1_____ YES 2_____ NO 3_____ UNDECIDED

26. If you have any additional comments about any question or item on this questionnaire, please write them in the space below or attach an additional sheet of paper. Thank you for your participation!

Appendix B



Roane County Department of Education

Jess H. Plemons, Superintendent

Telephone (615) 376-5592 • FAX (615) 376-1284

100 Bluff Road • Kingston, Tennessee 37763-9781

Roane County
Board of Education

Charles J. Crass
Chairman
622 Clinton Street
Harriman, Tennessee

Everett Massengill
Vice-Chairman
Route 4, Box 180
Kingston, Tennessee

Dean McDonald
Route 1, Box 155
Rockwood, Tennessee

C. R. Barger
Route 2, Box 299
Harriman, Tennessee

Troy Beets
2025 Cedar Lane
Kingston, Tennessee

Dr. Steve N. Smith
Box 620
Oliver Springs, Tennessee

Michael L. Miller
P. O. Box 523
Rockwood, Tennessee

Jane Harmon
Route 1, Box 191A
Harriman, Tennessee

Danny Wright
Route 1, Box 77
Harriman, Tennessee

August 25, 1993

Joseph B. Parker
1158 Circle Drive
Kingston, TN 37763

Dear Joe,

In regards to our earlier conversation, you have approval to utilize Roane County Schools' personnel for a preliminary survey for your doctoral dissertation concerning standardized testing.

Good luck in your endeavor and I wish you every success.

Jess H. Plemons
Superintendent, Roane County Schools

Appendix C

September 20, 1993

Dear Colleague:

As I am sure you are aware, the Basic Education Program for the State of Tennessee includes a component for teacher evaluation based on students' standardized achievement test scores. These scores are called "gain" scores and are a result of the teacher's, school's, and school district's effects on the student; thus leading to the term "value-added" assessment, the "value" being that added by the teacher, school, or school district.

Because the state is preparing to use standardized test results in this manner, I am interested in determining what perception of testing and terms associated with tests and measurements exists among teachers and administrators in the State of Tennessee. Therefore, I am conducting a study of those individuals and am asking you to become a participant by completing and returning the enclosed questionnaire. Your name was randomly selected by the Tennessee Education Association from a list of current Tennessee educators. The return of the questionnaire implies your informed consent to participate in the study. Your answers will be kept in total anonymity.

Please do the following:

1. Be as objective and as truthful as you can.
2. Answer all the questions on the questionnaire that pertain to you; please follow the directions.
3. Return the completed survey form in the enclosed, postage-paid envelope no later than Monday, November 1, 1993.

Your answers and comments will help me determine the perception of test awareness in the State of Tennessee. If you have any comments you would like to add, please feel free to write them on the the back. Thanks for your participation!

Joseph B. Parker
Doctoral Student, University of Tennessee

Appendix D

March 16, 1994

Dear Colleague:

In the fall of last year, you should have received a copy of the enclosed questionnaire about the use of standardized testing and your experiences with standardized tests in Tennessee. If you responded to the questionnaire, thank you very much; your input was very helpful. If you did respond, please check the box at the bottom of this page, tear it off, and return it in the enclosed self-addressed, postage-free envelope; no other action is required.

If you did not respond the first time, please take a few minutes to fill out the questionnaire and return it as soon as possible in the enclosed envelope. Your responses will help me determine a more comprehensive picture of testing and test awareness in the state of Tennessee. I need your input.

As I am sure you are aware, the Basic Education Program for the state of Tennessee includes a component for teacher evaluation based on students' standardized achievement test scores beginning in 1995. Because the state is preparing to use standardized test results in this manner, I am interested in determining what perception of testing and terms associated with tests and measurement exists among teachers and administrators in the state of Tennessee. Therefore, I am conducting a study of those individuals and have asked you to become a participant by completing and returning the enclosed questionnaire. Your name was randomly selected by the Tennessee Education Association from a list of current Tennessee educators. The return of the questionnaire implies your informed consent to participate in the study. Your answers will be kept in total anonymity.

Your answers and comments will help me determine the perception of test awareness in the state of Tennessee. If you have any comments you would like to add, please feel free to write them on the back of this letter or on the questionnaire. Thanks for your participation!

Joseph B. Parker
Doctoral Student, University of Tennessee

Please detach and return only if you responded to the initial questionnaire!

☐ I responded to the initial questionnaire. I understand that
my responses will be kept in total anonymity.

Appendix E

Responses to Open-Ended Question #26 or Other Comments:

"If all students came to testing site with no worries or emotional problems I would agree. We need to take outside factors (such as home problems or physical problems) into consideration."

"I do not feel that teachers should be evaluated by achievement test scores. I do not believe one score by a child is a valid indication of what the child can do or has learned. I do not think it is valid for the class scores of a teacher to be compared with those of the class the year before."

"I teach honors English to college bound students and I must admit I am more interested in their scores on college entrance tests than I am on TCAP. I am much more interested in preparing them for college composition classes than I am in their TCAP scores."

"Too many variables that affect students' learning and performance will never be taken into account through standardized [sic] testing. Therefore, basing a teacher's performance on just these results is ludicrous."

"I think judging a student on his or her test scores is one of the most unfair ways to judge a student. Please let the teacher go back to grading a student to the best of our abilities. Boo to test scores."

"Most kids at the school I teach just mark answers and could care less about there [sic] scores. If teachers are evaluated on this criteria there will be alot [sic] of kids fail to hurt the teacher on purpose."

"In H.S. science most important material (lab problem solving, field techniques and problem solving, observation and creative thinking) can't be tested on a paper test. To good sci. ed. standard tests are usually a waste of time and paper."

"I fear that teachers will find it necessary to desire to cheat by copying tests, changing student answers, etc. in order to produce desired results. I have seen teachers copy tests and put charts up in a classroom which told answers."

"I work with 3-5 yr. old autistic children and have never given or used standardized tests."

"The test system is as screwed up as our State Government. Thats [sic] bad my friend!"

"Until testing is done professionally and all follow the same rules--test scores mean very little. I see too many high test scores, but students can't perform, therefore

teacher evaluation based on test results will never work. TCAP's is [sic] a poor instrument and should be discontinued."

"To evaluate a teacher's performance on some test made in California and adapted to Tennessee is something you would hope came from some production company. If teachers' evaluations are going to be based on the junk called tests we use now, we're in trouble. Get out of the ivory towers of administration and see what education really is! We deal with warm bodies and not politics and red tape. Wake up and smell the coffee! We are teaching! Administration has its place. But, we are the teachers! They aren't!"

"In regard to question 10: My responses are based on tests such as the ACT, PSAT, SAT, DAT. Certain tests address some skills more than others. For instance, the DAT doesn't test higher-order thinking to the degree that the others do. I'm not sure my answers are valid."

"Poor way to evaluate teachers."

"I teach Chapter I students therefore I rely on TCAP or other standardized test scores to determine eligibility, curriculum used and skills taught."

"I teach first grade and I think testing young children is ridiculous! A standardized test tests a child on those particular questions on that one day of the year. It has nothing to do with their 'achievement' or teacher effectiveness!"

"There are so many variables that we can't control, especially things such as the home environment. It would be foolish to assume that teacher evaluation based on test scores will be a cure-all for problems faced in the classroom."

"How does the evaluator know? How do you know that a student is giving a 100% on a test? How do you make a student care if he does well when there is no encouragement [sic] from home? If a child is not feeling well, how is that noted? Some students are lucky guessers."

"Every group of students is different. The teacher and the school have no control over this. Our state evaluation model is comprehensive enough to assess teachers."

"I teach history, and because the texts are so narrow, biased and full of errors, I use many other sources so my students will learn the truth. The TCAP rarely matches the basic curriculum and certainly does not measure knowledge of a broad spectrum and totally avoids any higher order thinking skills. For example, on last year's TCAP, the Social Studies section had fewer than twenty questions, and only four were even related to my grades's curriculum, and those four questions were about minor, insignificant details which were meaningless.

"To use the TCAP to measure student or teacher performance is another example of non-educators (bureaucrats, politicians, statisticians, etc.) using education to further their own careers while continuing to ignore the real reasons our students are not learning better (overcrowded classrooms, no materials, inadequate materials, etc.)

"Until those in control of education (the non-educators) begin to address real

problems and stop using the schools as a toy, things will continue to stagnate."

"Students can not feel well when taking the test, scores will not be as good. In K-5 test should be write [sic] in booklet not transfer answers to test form."

"I am a strong opponent of teacher evaluation based on student test scores. I feel that it would greatly inhibit effective, creative teaching."

"Although we look at, consider, discuss these standardized test scores, I feel that they do not fairly reflect any student's or teacher's efforts. The tests that are 'forced' on us do not accurately represent the students we teach nor do they follow state curricula guides--they test nothing we are supposed to teach in Tennessee."

Pertaining to question 10: "Don't often provide usable info that can be used to benefit the kids."

Pertaining to question 11f: "Especially not reflective if child functions out of test's range but is required to take test 'if he can hold a pencil.' ie--functioning on 2nd grade in reading level but required to take 8th grade test. This is ridiculous. Not valid, reliable, or usable. Should be allowed to take out of level test (TCAP)."

Pertaining to question 10c: "Not the same for every child."

Pertaining to question 11f: "Not necessarily true--some students just do not do well on standardized tests."

"A lot of students do not take standardized tests seriously. Many of them just mark answers to get the test finished.

"Too much emphasis is placed on test scores rather than learning."

"I strongly disagree with standardized test being a tool for evaluating teacher performances because there are many factors to consider:

1. Some students don't perform well in class due [sic] excellent on test.
2. Some students that perform well in class are not good test applicants.
3. Some of the good test applicants could have a bad day and not apply their best performance on a test. I am speaking from experience because I've had this to happen with 'good' students.
4. There are A + B students that panic when they are informed, this is a standardized test.
5. Speaking from my own personal experiences, I become very nervous during standardized test. There is a phobia that I am being timed, I must rush, yet in classroom setting one is told to take your time and do a good job.

"If teachers are going to be evaluated on standardized test, then the methods of teaching must change to correspond with the methods of taking test. Example: Students should be timed on all classroom/subject test given in a normal school day. Papers should taken when the 20 or 30 minutes is up.

"All classroom/subject test should be computer answer cards to bubble.

"All classroom/subject test should be given on numerous subjects within a given day.

"The standardized test is given within a maximum of three days and all subject

areas are tested.

"All the textbooks and commerical [sic] tests should be re-designed to be administered just as the standardized test.

"If all these adjustments are made, I can accept such an evaluation.

"The methods of teaching and testing used in the classrooms now is contradictive to a valid evaluation of standardized test.

"The curriculum is designed with designated time periods for various skills to be taught and the standardized test are given in April before all skills have been taught. This is very unfair.

"Example: I find my children score lower on the geometry section of the Math because it isn't schedule [sic] to be taught until after the standardized test has been given. There are other areas and skills that aren't scheduled to be taught until after the test is given."

"I am opposed to evaluating teacher performance due to test scores. If you could see some of the students that teachers have to deal with, you would understand why."

Pertaining to question 22: "Supervisors talk to each other. They need to go back to the classroom--principals, too.

"Until parents and the business world become more concerned and insist that students perform at a higher level, all the testing in the world won't really matter."

"When drugs, abuse, and sleepless nights can be controlled in a child's life, then and only then should teachers be judged on their test results. We do the best we can and get little or no respect."

"Can we be held accountable for students who sleep in class and have little or no self-motivation or parent involvement? I don't think we can fairly be judged by TCAP scores."

"I have had several students transfer into my classroom right before TCAP and there is no way I can make sure they are totally prepared for the test."

"Pertaining to questions 10 and 11: "TCAP test scores are not an accurate reflection of student achievement or teacher performance because factors such as student motivation and home environment are not taken into account."

Pertaining to #24: "I think using students test results as a component of teacher evaluation is unfair and invalid to a teachers ability or teaching methods."

"The teachers I hear on this subject are against using these scores for teacher evaluation. I haven't given it a lot of thought, since it hasn't targeted the high school yet."

"I teach elementary physical education and do not use standardized test scores except in the fitness tests that have been given. I usually refer to standardized tests when checking on a specific student who may be having difficulties. I hope this doesn't throw your research off. I do remember from Ed. Research the importance of

getting questionnaires returned."

"Any testing program should go to a similar format to the AP exams but with other things added. Every sub-test part should have short answer question (mult. choice); student generated answers (new on PSAT math); writing responds (AP format); and reading and responding in that sub-test area.

"Teacher evaluation based on standardized testing is bull manure. Present tests are not good enough, do not measure the right things, and are not in the proper format to measure a 'real' education. When (if????) we can make a 'proper' test; it's [sic] main purpose should be 'student' accountability not teacher accountability (and absolutely not teacher evaluation). We have tried educational reform based on everyone's accountability except the student. We test the student, then try to lay blame for poor results on everyone and everything except the student. The only educational reform that is going to have meaningful results is the one where student accountability is the 'main' part. Some ideas for student accountability longer classes (2 hours for 1 hour credit), longer day (mandatory before and/or after school sessions), longer year (mandatory summer sessions), Saturday sessions (mandatory), mandatory homework (no homework = no grade), no extra school activities without standards (sports or clubs; all grades C or above; all homework; all standards; attendance) and no credit for class till all standards are met (attendance, 'every' piece of work is completed to a C or better standard, student portfolio done, a 'good' standardized test passed in subject area . . .)"

"Student's performances/abilities are not always accurately reflected by TCAP scores. All testing is not equal in all school districts. One district nearby had teachers who boasted of seeing tests the day before students were tested. Students were given answers to the test at that time. Why not use national tests? Because jobs have been created to develop the TCAP. Teacher uprising is on the way!"

"Student test scores are not very valid in my area due to many factors not controlled by system. All test scores on human performance should not determine the effectiveness of the teacher. There are too many outside factors involved!!!!!!!"

"I personally feel that tests should be designed by using material from the existing curriculum and with higher-order thinking questions should be submitted by the school/county administrators and teachers."

"Teachers should be held accountable for student's test scores when teachers carry those students in utero and when those students go home with those teachers every night."

"I think looking at individual student gains over one year to the next would yield useful information regarding individual student's progress, but I'm not convinced that it's an effective way to evaluate teachers. I do believe this will do a great deal to achieve uniform curriculum across the state by grade level."

"The use of standardized test scores to evaluate teachers is unfair. Students should never be allowed to social[ly] pass grades. Each grade should have a standard of passing. There should be special school rooms and classes in each school for students

who cannot pass. Have a class on test taking skills, test making, etc."

"There are too many other factors involved in student learning to evaluate teachers according to standardized scores. Often tests do not evaluate learning."

"I returned this before -- I teach elementary and middle school art and don't use test scores in any way -- I don't administer tests or have any contact with them. I do feel it's totally unfair to teachers to judge them on 1 set of test scores given in a year. There are too many factors that determine a teacher's success with a certain group of students. Only if tests were given at the beginning and end of each school year could a halfway fair assessment of a teacher be made."

"Until scores are readily available to classroom teachers -- little use -- Now I have to go to guidance and search file. Also, need a periodic workshop regarding statistical terms, etc., if we are going to communicate using such info."

"Standardized tests require an entirely different method of teaching as opposed to teacher-made tests: Test must be reviewed by teacher before teaching chapter/unit to make sure all components are covered. Don't care for standardized tests. Teacher-made are much more effective."

"This method of teacher evaluation will not measure enough skills in the arts to warrant its use as a tool. This is a poor substitute for administrative evaluation."

"Thanks for allowing me to complete your questionnaire concerning the use of the T-CAP and other standardized tests that are currently being used in schools all over the state of Tennessee. These tests are good for diagnostic purposes, such as using the scores to determine students who need special services. However, tests should not be the only criteria used. Teacher judgment in most cases is more valid than a standardized test."

"Concerning the legislator's decision to use the results of the test to evaluate teachers, I feel this is another example of politicians drawing conclusions about a subject without personal knowledge. I do not believe that a teacher or administrator should be evaluated on the basis of one standardized test. There are several reasons why I believe this:

1. Test scores are not always valid.
2. Students may be ill during the testing situation.
3. Students with special problems, such as ADD, may not have taken their medication.
4. Students may have had an argument with their parents before coming to school.
5. Some students are poor test takers. Their daily work, homework and oral answers are average or above. But the test does not show this.
6. Some students become very nervous when they are being timed on a test. Thus affecting their concentration and performance.
7. Some students work very methodically and therefore are unable to finish the test in the allotted time. The unanswered questions are counted as wrong. This gives an invalid score of this child's achievement.
8. All students are not the same. They have different abilities, strengths, interests, home environments, and levels of motivation. Schools can not be conducted as

a business, putting in the raw material and turning out a perfect and exact product.

9. According the State Testing Committee, the T-CAP test was never meant to be a Value-added test.

10. Legislators have been invited to visit the schools in our system as I am sure they have in many others. Very few if any have taken the time to visit, observe, or ask teachers for their input. Therefore, I do not think that they are qualified to evaluate our performance as a teacher.

"There are also parent reactions to consider. If the test scores are printed in the newspaper, as we have been told, the parents will naturally request the teacher whose class scored the highest on the test. The teacher that scored the highest may have been lucky enough to have an above-average class as compared to another class that consisted of mostly Resource or Chapter I students. The diverse level of achievement found in regular classrooms makes it impossible to evaluate the teacher based on how her students performed on the T-CAP test.

"I hope I have been of some help. As you can see I feel very strongly about the topic of teacher evaluation based on test scores. Teachers should be evaluated, but by someone who is familiar with education and can make an objective decision. At this time, teachers are evaluated by the SACS association. We, as teachers, are trying desperately to teach each and every student and help them to become a productive adult."

"I am a Special Ed. teacher of L.D. and mildly retarded students. They are not required to take standardized tests.

"Sorry I could not be of help to you." (left entire instrument blank)

"I have not been involved with tests for some time. I prefer not to complete this questionnaire."

"There are far too many aspects of the children and their lives that cannot be accounted for on a standardized test. There is NO TEST that can account for what that child may be feeling or thinking during the testing period!!!!!!! One test proves nothing . . ."

"I'd hate to think my purpose for teaching is only for the purpose of testing and the above terms mentioned but sadly it seems it's coming to that. What the students do or how they perform on one 'given' day to one given test -- T-CAP."

"I think it is absolutely ridiculous to base the results of testing to determine the quality of teaching. In my experience I literally beat my brains out trying to teach skills according to state mandation (Tenn. Curriculum Guide) but all students do not respond. As the old saying goes, 'You can lead a horse to water but you can't make it drink.' If it doesn't then, is it my fault? I can't drown it trying! Whoever devised this plan should try teaching a while."

"I feel that to evaluate a teacher's performance based on a test modified to sell to a particular state is rather unprofessional. This will lead to more test 'teaching,' tampering with scores, and out and out cheating! Students', parents', and the publics' attitude are not considered. What if they don't care? (As is the case much of the time). The opinion of education is not if students learn, but rather if they can perform well on a standardized test!!"

"After teaching all day, I don't know if I comprehended some questions correctly or accurately. You should try to teach 6th graders! Problems! Problems! Problems!"

"I believe it is unfair to base a teacher's credibility on standarized [sic] testing. Although a student may make A's and B's in class, they may score low on standarized test [sic]. That is not a reflection of poor teaching, but shows a student who is a poor test taker."

"I wrote this poem several years ago while administering the Stanford tests to my freshmen classes. I thought you might be interested:

ON TESTING

by J. Earl Stroup

"Measure what your students know.
Give them a test, and it will show
All of the things that they do best.
Here is the answer--Give them a test!

"Measure what your students learn.
It matters not how much they yearn
To see a robin, or touch its nest.
Do they know biology? Give them a test.

"Measure who your students are.
Testing's a better method by far
Than caring, love, and tenderness.
Find the answer: Give them a test.

"Alcohol, drugs, teen suicide,
Headaches, ulcers, pains inside,
What has caused this undue stress?
Could it be: just too many tests?"

"Standardized test results are not a vailid [sic] or reliable means of assessing any teacher's performance, nor should they ever be a component of teacher evaluations. Teachers can not make a child do his/her best. The manner in which a lot of items are stated, the terms used, etc., the way problems (in math) are written causes children to miss them, even when they know how to do them."

"There is little in the communications field that is directly tested by the standardized tests we use, it seems to me.

"ESL students who are beginners in English, as my students are, are usually unable to take the standardized tests."

"I do not find the TCAP to be a valid or useful tool. By useful I also mean that there is nothing specific enough returned to us to use diagnostically. The scores within

the test often conflict. The insistence and inflexibility to give only grade level tests as opposed to functioning level tests is contrary to anything I've learned in testing and yields totally invalid scores. (A student who can only read at a 2nd grade level but must take an 8th grade test) Some of these students are never certified for special ed. and are used for value-added. RIDICULOUS!"

"I think too much emphasis is placed on these tests. Some students think school is over after tests are taken. They get this idea from teachers. It makes you want to teach only for the tests as some schools do just this.

"I don't believe this test can accurately evaluate a teacher's teaching ability. Too much is determined by students' attitudes toward the tests themselves. Some students do not have the encouragement from home or the desire to do their 'best'."

"If a nation could be tested into excellence"

"Standardized tests do not always adequately measure students' abilities or inabilities. Some individuals are comfortable taking a standardized test while others cannot function on this type of evaluation."

"I have poor results from student test scores. The students do not perform as well as scores indicate. Teachers have told students answers to questions to improve test scores causing invalid results. Students think achievement tests are a 'joke' and do not take it [sic] seriously. Students finish a test in less time than it takes me to do the test!! Students coming to me can not be evaluated[sic] by test scores from previous achievement tests. Teacher evaluation by test scores will result in SUPER invalid results."

"Teachers in nearby counties boast of seeing the T-CAP test the day before the T-CAP is given. Then they teach the students the info found on that entire day before the test. Evidently, this behavior was condoned by principals and others at the schools where these incidents happen. Students in many districts are being encouraged by parents to do poorly on work and misbehave in order to receive SSI benefits. Please do use the results to forward to the Governor and others. Teachers can not and should not be held responsible for students' test scores.

"The test scores are also unfair to students. A tremendous amount of stigma is placed upon students who do poorly on tests. Some of these same students may achieve otherwise.

"Sometimes the T-CAP tests uses [sic] questions that that are far beyond the grade level of the students. When there are very few questions, then the students become frustrated and do poorly. How can subjects such as science be accurately tested when there are only 15-20 questions? The same goes for other courses!

"Students have voiced their concern over having other teachers administer the test. The students prefer to be in their own classroom with their own teacher.

"Parents are happy with the tests until their child does poorly!

"Also, why does Tenn. have to use it's [sic] own testing instrument? Is the state employing relatives to those in office to construct and score the tests? There seems to be a lot of wasted money in Tenn. that could be used more constructively! Parents and other educators have asked this question frequently."

"I am a foreign language teacher who also teaches English occasionally. I have found that only the TN Proficiency test is used at my school to place kids in a lower level English class. Other subjects, (except math) are not affected."

"I seem to give negative comments about standardized tests being used to evaluate my effectiveness in class. If the kids were to be tracked not only with scores but also with outside factors considered, I might be more open-minded. If a child has high absenteeism, how can you expect a teacher to be held accountable when the parents are not! Who's responsibility is raising the child anyway? We no longer just teach academics--we are expected to make them acceptable in society when some parents do nothing!"

"Each year something happens to make me feel stronger about the unfairness of using test scores. I have not talked to a single teacher that feels these scores are valid according to a child's actual progress. Two weeks before test dates, I enrolled a new student in my room. After an hour I realized this child knew nothing. He couldn't read a word and had not heard of addition and subtraction. After a week had passed I received his records and discovered he had not even been in 1st grade. It was decided [sic] I would go ahead and keep him in my room and just retain him in 1st grade. I went ahead and gave this child the same test I was giving the rest of my 1st graders and this is when I really hate achievement tests. My average students were really having to struggle to try and read a test way too hard for them and were doing their bests. I stood and watched my new student bubble an answer for every box (not able to read a single word) and he was getting about 1/2 right. Do not try and tell me this test will measure this child's progress for he was only a lucky guesser."

"Dear Sir, You have asked for my input on the subject of value added assessment. I currently teach 6th grade social studies. In my curriculum guide, I have forty plus skills that my students should be able to do by the end of the year.

"I see problems developing with this new plan when:

"#1 The test is given two months before the end of school.

"#2 The social studies part of the T-CAP only has 20 questions.

"#3 The majority of questions asked are related to textbooks and subject material taught before my grade level."

"I strongly feel that using the TCAP scores as an evaluation of a teacher's performance is ineffective, inconclusive, and extremely unfair. The reasons for this statement are based on the fact that since teachers cannot control the quality of the product (students) from beginning to end, they should not be held accountable for the finished product. There are too many variables which influence the student's performance to hold one segment of their development responsible. Having watched students being tested for over twenty years, I can attest to the fact that their scores will be just as adversely affected by their mood that day, by the quality of their homelife, by their parents' attitude about education, by their lack of self-motivation, and by their health as it is by the quality of their instruction. Therefore, to use these scores as a measure of their teacher's instructional quality is neither realistic nor productive."

"Most teachers are against value added assessment."

"I teach visual art and rarely have a need or interest to look at standardized test results."

"More interested in developing authentic/alternative assessment tools/rubrics/systems than in standardized [sic] testing! TCAP is a dinosaur! Plus, it does NOT even follow our state curriculum -- tests on subjects not even in course of year's study!! Teach A, test B. Stupid."

"I certainly do believe in the value of standardized testing when the student knows it counts for him/her. (ACT, SAT, etc.) I believe the notion that the student who does poorly or shows no gain (even though that individual may have closed at McDonald's the previous night in order to earn money for car insurance . . .) makes a major statement about the effectiveness of his/her teacher is absurd!"

"The T-CAP is ludicrous! Value added stinks!"

"I strongly feel that T-CAP scores are not an accurate means of assessing teacher performances. There are too many other factors which make them inaccurate."

"Too much emphasis is placed on T-CAPs. Many teachers are threatened and students feel the pressure to do well from home and school. Students dread T-CAPs and are uptight -- so do teachers! Many are not good test-takers."

"If test scores become a way of evaluating teachers, then they will become less valid in evaluating students. Some teachers will resort to any method to try to make themselves look successful."

"To me the first teachers (the parents) play an extremely important role in how the child performs on a test. Some parent [sic] tell them don't worry about how you do, it (TCAP) is just to judge the teachers not you. The media gives that impression."

"I definitely do not feel achievement tests are an accurate measurement of a teacher's or student's performance. One test can not give a true picture for a year of hard work. One bad day of testing can result in poor results when in reality the teacher covered the objectives and the child mastered, but did not feel well, nervous, etc. I feel if we want our children to improve we need to slow out curriculum down. Our children are expected to do too much!"

"The T-CAP test seems on certain items that it uses vocabulary to trick students for a skill. For example: using a geography listing of an area that students are not familiar with to test capitalization. If a child understands the skill, why not use a city, country, etc. that's more familiar to see if they know the skill, not the vocabulary."

"90% of my classes is [sic] in physical education, its [sic] hard to use tests to grade most of students. A grade is given on individual performances."

"To me, this is no way to evaluate teacher's performance. I know I do my job to

"To me, this is no way to evaluate teacher's performance. I know I do my job to the best of my ability, but not all students even try on the test!"

Vita

Joseph B. Parker was born in Atlanta, Georgia on December 22, 1946. He attended the Atlanta City Schools and was graduated from Marion Military Institute, Marion, Alabama, in 1965. He enlisted in the U. S. Navy in 1966 and attained the rank of PO1 (E-6). During his military service, his fields were electronic communications and reactor control systems. He made two tours to Vietnam and was honorably discharged in 1975.

He attended Roane State Community College and the University of Tennessee and was graduated with high honors in 1979 with a Bachelor of Science in Education. He began his professional career by teaching eleventh and twelfth grade English at Loudon High School, Loudon, Tennessee, in 1979. He was to remain there for three years as an English teacher until moving to Cherokee Middle School in Kingston, Tennessee, in 1982. Here he taught seventh and eighth grade English and served as co-sponsor of the Junior Beta Club. In 1985, he transferred to Roane County High School as both an English and psychology teacher. While at the high school, he was the sponsor of the Student Council for three years, the Annual sponsor and journalism teacher, a member of the Attendance Committee, and Chairman of the School-wide Calendar Committee.

In 1983, he was awarded a Master of Science in English Education from the University of Tennessee. In 1989 he returned to the University to pursue a doctorate in Educational Administration and Supervision with a collateral area in Management. He has been a member of the National Council of Teachers of English from 1979 until 1988, and is currently a member of the National Association for Supervision and Curriculum Development, the Tennessee Association for Supervision and Curriculum Development, the International Society for Technology in Education, and the East Tennessee Society for Technology in Education where he serves on the Board of Directors from Region 2. In the spring of 1991, he was inducted into the Phi Delta Kappa honor society.

Currently, he is Technology Coordinator for the Roane County Schools. He is married and lives with his wife, Patricia, on the shores of Watts Bar Lake in Kingston.