PSYCHOLOGICAL AND EDUCATIONAL TEST SCORES

PSYCHOLOGICAL AND EDUCATIONAL TEST SCORES

What Are They?

By

MARTY SAPP, ED.D.
Professor
Department of Educational Psychology, Counseling Area
University of Wisconsin–Milwaukee

Published and Distributed Throughout the World by

CHARLES C THOMAS • PUBLISHER, LTD.
2600 South First Street
Springfield, Illinois 62704

ISBN 0-398-07258-2 (hard)
ISBN 0-398-07259-0 (paper)

Library of Congress Catalog Card Number: 2001043369

With THOMAS BOOKS *careful attention is given to all details of manufacturing and design. It is the Publisher's desire to present books that are satisfactory as to their physical qualities and artistic possibilities and appropriate for their particular use.* THOMAS BOOKS *will be true to those laws of quality that assure a good name and good will.*

Printed in the United States of America
SM-R-3

Library of Congress Cataloging-in-Publication Data

Sapp, Marty, 1958-
Psychological and educational test scores : what are they? / by Marty Sapp.
p. cm.
Includes bibliographical references and index.
ISBN 0-398-07258-2 -- ISBN 0-398-07259-0 (pbk.)
1. Psychological tests. 2. Educational tests and measurements. I. Title.

BF176 .S26 2002
371.21--dc21

2001043369

To the students on my research team.

PREFACE

Psychological and Educational Test Scores: What Are They? is designed for individuals who are forced to interpret test results or scores–regardless of their understanding of tests–such as psychiatrists, social workers, school principals, admissions personnel, physicians, attorneys, teachers, and parents. Because test scores can be easily calculated with a calculator, such as the Texas Instrument TI-30xa, or a computer program, such as SPSS, SAS, and EQS, this text is designed to give readers the confidence to understand and calculate test scores; therefore, with a few key strokes of a calculator, or a few commands on software, you will easily be shown how to calculate various test scores.

With over 13 years of experience working with a variety of master's level and doctoral students within Educational Psychology and Counseling Psychology, I have found that many students, even those with an elementary or intermediate training within psychological and educational test theory, lack the ability to interpret Z-scores, t-scores, MMPI–2 scores, grade-equivalent scores, and so on. Furthermore, I have found that many students and professionals within the areas of psychology and education are unaware that modern psychological and educational tests are based on three dominant psychological test theories: true score theory, item-response theory, and generalizability theory. Moreover, I have found that many users of psychological and educational test results need a narrative understanding of these theories, and they need to understand how these theories influence test scores and other factors, such as reliability and validity.

For the nonquantitatively-oriented reader, there are good advanced texts within the area–*Psychological Testing* (7th Edition) and *Psychological Testing: Principals, Applications, and Issues* (5th Edition); however, I have found that the reading level, mathematical ability level, and extensive topic coverage are more than the average layperson or nonquantitatively-oriented reader can assimilate. In addition, neither text is designed for the quantitatively phobic reader. Finally, neither text connects test scores with computer software or shows the reader how to perform calculations via calculators.

If you are a nonquantitatively-oriented reader or part of a lay audience, in terms of test applications, you need to know what Z-scores, t-scores, IQ scores, stanine scores, and percentile scores are, and you need a narrative description of the factors that affect reliability and validity. For example, t-scores, common psychological or educational test scores, are difficult to understand within their abstract forms, but you will find that you can get an intuitive feel for t-scores once they connect with psychological tests such as the MMPI–2, State-Trait Anxiety Inventory (STAI), Test Anxiety Inventory (TAI). Moreover, you will find that IQ scores are more meaningful when they are connected with real-life examples applied to the Wechsler intelligence tests.

In conclusion, a text that addresses psychological and educational test scores for lay audiences and nonquantitatively-oriented readers has not heretofore been available, especially a brief introduction to the practical applications of test scores. Essentially, this is an overview text to psychological and educational testing for nonpsychologists and nonpsychology majors; however, a basic understanding of the topics covered within this book will help anyone to understand what psychological and educational tests can and cannot do. Finally, this book is designed for the nonquantitatively-oriented users of psychological and educational tests, and it links test scores with computer software and the worldwide web.

ACKNOWLEDGMENTS

It took several individuals to bring this text into press. First, I would like to thank the students on my research team. Second, I offer thanks to Ms. Cathy Mae Nelson and the University of Wisconsin–Milwaukee School of Education word processing pool for typing this entire manuscript. Moreover, I thank Dr. Walter Farrell at the University of North Carolina at Chapel Hill. In addition, I offer thanks to my University of Cincinnati connections: Dr. Patricia O'Reilly, Dr. Judith Frankel, Dr. Marvin Berlowitz, Dr. Purcell Taylor, Dr. David L. Johnson, and Dr. James Stevens. I offer special thanks to Dr. Stevens, because he taught me how to embrace the quantitative aspects of being a psychologist. Finally, thanks also go to Dr. Cheryl L. Johnson at Miami University of Ohio. In closing, comments or discussions concerning this text–both positive and negative–are encouraged. My address is The University of Wisconsin–Milwaukee, Department of Educational Psychology, 2400 E. Hartford Avenue, Milwaukee, Wisconsin 53211. My telephone number is (414) 229–6247, my e-mail address is sapp@uwm.edu, and my fax number is (414) 229–4939.

CONTENTS

Page

Preface vii

Chapter

1. BASIC STATISTICS 3
   - Steps in Test Construction 3
   - Levels of Measurement 3
   - Frequency Distributions for Test Scores 4
     - Proportions and Percentages 6
     - Cumulative Percentages 8
     - Interpolation Example 9
     - Measures of Central Tendency 10
     - Using the TI-30xa to Calculate the Mean and Standard Deviation 11
     - Measures of Variability 11
     - Standard Scores 12
     - Calculating Measures of Central Tendency and Measures of Variability Using SPSS and SAS Computer Packages 15
     - Correlations 16
     - Four Covariance Rules 17
     - Correlation Exercises 20
   - Chapter Summary 21

2. THEORIES OF PSYCHOLOGICAL AND EDUCATIONAL TEST SCORES 22
   - Classical True-Score Theory 22
     - Seven Assumptions of the True-Score Theory 23
   - Generalizability Theory 26
     - Sums of Squares for Individuals 28

SPSS Control Lines for Internal Consistency Reliability Analysis . . . 31
Reliability Coefficients within Generalizability Theory . . . 32
Summary . . . 32
Item Response Theory (IRT) . . . 33
Chapter Summary . . . 36

3. RELIABILITY . . . 37
Reliability . . . 37
Types of Reliability . . . 38
Factors Affecting Coefficient Alpha and Other Forms of Reliability . . . 44
Reliability of Differences Scores . . . 46
Reliability of Speeded Tests . . . 48
SPSS and Reliability: A Practical Example . . . 48
Chapter Summary . . . 60

4. VALIDITY . . . 62
Chapter Overview . . . 62
Face Validity . . . 62
Content Validity . . . 63
Criterion Validity . . . 64
Predictive Validity . . . 64
Construct Validity . . . 66
Exploratory Factor Analysis Using SPSS . . . 68
SPSS Commands for Factor Analysis . . . 76
Confirmatory Factor Analysis or Structural Equation Modeling Using EQS . . . 79
Explanations of EQS Commands . . . 82
Selected Output . . . 82
EQS Commands . . . 86
Selected Output . . . 86
Standard Error of Estimate . . . 92
Factors Affecting Criterion Validity . . . 93
Decision Theory and Test Items' Validity . . . 93
Nonequivalent Control Group Design . . . 101
Meta-Analysis . . . 102
Chapter Summary . . . 108

5. PULLING IT TOGETHER 109
Reliability/Validity of Test Scores 109
Factors that Can Affect Correlation, Reliability, and Validity Measures 110
Chapter Summary 111

6. COMMON PSYCHOLOGICAL AND EDUCATIONAL TESTS 112
Chapter Overview 112
Minnesota Multiphasic Personality Inventory–2 (MMPI–2) 112
Basic Scales of the MMPI–2 113
Profile Invalidity 113
Clinical Scales 115
MMPI–2 Code types 116
Section Summary 120
Millon Clinical Multiaxial Inventory (MCMI) 121
California Psychological Inventory (CPI) 121
Edwards Personal Preference Schedule (EPPS) 122
The Strong Interest Inventories 123
Tests of Mental Ability 124
Wechsler Scales 124
Stanford-Binet Intelligence Scale 127
Woodcock-Johnson Psychoeducational Battery 128
Wide Range Achievement Test–3 (WRAT–3) 128
Group Ability Tests 129
American College Test (ACT) 129
Scholastic Assessment Test (SAT) 130
Graduate Record Examination (GRE) 130
Miller Analogies Test (MAT) 131
Projective Techniques 131
The Rorschach Inkblot Test 131
The Holtzman Inkblot Technique 132
Thematic Apperception Test 132
Draw-A-Person Test (DAP) 133
Depression and Anxiety Assessments 133
Beck Depression Inventory (BDI) 133
State-Trait Anxiety Inventory (STAI) 134
State-Trait Anger Expression Inventory (STAXI) 134
Test Anxiety Inventory (TAI) 134
Chapter Summary 135

7. INTERPRETING PSYCHOLOGICAL AND EDUCATIONAL TEST SCORES WITHIN A PSYCHOLOGICAL REPORT 136
Chapter Overview 136
Sections of a Psychological Report 136
Psychological Report 1 137
Commentary 138
Psychological Report 2 139
Commentary 141
Psychological Report 3 141
Commentary 148
Psychological Report 4 149
Commentary 155
Psychological Report 5 155
Commentary 158
Psychological Report 6 159
Commentary 164
Chapter Summary 164

8. PSYCHOLOGICAL TEST SCORE TERMS 165
Chapter Overview 165
Objective Tests Versus Subjective Tests 165
Free Response Items Versus Recognition Items 165
Speed Tests Versus Power Tests 166
Dichotomous Items Versus Multipoint Items 166
Maximum Performance Tests Versus Typical Performance Tests . 166
Normative Measures Versus Ipsative Measures 166
Forced-Choice Tests 166
Response Biases 167
Response Styles 167
Scaling Models 167
Probability, Monotonic 168
Deterministic, Monotonic 168
Probability, Nonmonotonic 168
Item Analysis 169
Chapter Summary 170

References 171
Author Index 179
Subject Index 182

PSYCHOLOGICAL AND EDUCATIONAL TEST SCORES

Chapter 1

BASIC STATISTICS

STEPS IN TEST CONSTRUCTION

First, a **psychological** or **educational test** is a systematic and an objective measure for a sample of behavior. And **measurement** is the assignment of numbers to test scores based upon **rules**. Essentially, there are **six steps** to **test construction**. **First**, a **table of specification** is developed that specifies the **purpose** of a test. **Second**, from the table of specification, **the form and content of test items are generated**. **Third**, items are tried out with individuals and modified based on feedback from the individuals taking the test and the individuals administering the test. **Fourth**, the procedures for administering and scoring the test are standardized. **Fifth**, developing **reliability** (consistency in the measurement of test scores) and **validity** (determining if the test items are measuring what they purport to measure) data for the test. **Sixth**, developing norms (average or standards of test performance) for the test (Anastasi & Urbina, 1997; Sapp, 1997). Returning to the notion of measurement, the next section discusses levels of measurement.

LEVELS OF MEASUREMENT

One aspect of psychological and educational test scores that is confusing is **levels of measurement**. Psychologists have broken psychological and educational test scores down into four levels or hierarchies–**naming**, **ranking**, **quasi-interval**, and **real intervals**.

Working from the lower level upward, **nominal levels of measurement** is just simply the **naming** of variables. For example, the names in a telephone book represent nominal levels of measurement (Adams...

Banks...Cable...Zacka). Nominal levels of measurement represent **distinctness** or naming, but not ranking, quasi-interval, nor real interval scores.

Ordinal measures assign numbers to represent the **ranking** of some distinct characteristic. For example, ordinal measurements are used by judges to rank the distinct characteristics of beauty for Miss America and Miss Universe pageants. In terms of the hierarchy, ordinal forms of measurement include the naming and ranking of characteristics. **Quasi-intervals** are common levels of measurement within psychological and educational testing, and they include scores from many tests, such as IQ scores from IQ tests, achievement tests, personality tests, such as the MMPI–2, and college entrance examination tests. Quasi-intervals are obtained when equal differences among measurements represent the same or equal amount of difference. For example, the difference among 80° F (Fahrenheit), 90° F, and 100° F is an equal distance of 10° F; however, the difficulty with quasi-interval scores, such as IQ scores, is the difference between an IQ score of 80 and 90 and does not have the same meaning as the difference between IQ scores of 120 and 130.

Furthermore, quasi-interval scores do not have **real** or **absolute zero points**; for example, what is the meaning of zero IQ? Or does a temperature of 0° F represent zero or the absence of temperature? Clearly, zero is a relative value for quasi-interval test scores.

Ratio scales of measurement have **"real"** intervals. For example, 6 inches is twice the distance or length of 3 inches. And zero inches represents that no inches exist; therefore, ratio scales, many of which are scales used in physics and chemistry, have absolute or real zeroes. With ratio scales of measurement, zero is meaningful. And ratio scales have the properties of quasi intervals, ordinal and nominal scales of measurement.

FREQUENCY DISTRIBUTIONS FOR TEST SCORES

A frequency distribution is a method of tallying data collected on a population or sample, and it is a method for representing test scores. Frequency distributions are usually performed on populations and samples. A population is all kinds of people or things that are being studied. Similarly, a sample is a subset of a population, and if it is obtained in a random manner (each datum has an equal chance of being chosen), inferences can be made from the sample back to the population.

Suppose we gave a 9-word spelling test to 100 middle school students, and 2 students (frequency) spelled all 9 words correctly, 5 students (frequency) spelled 8 words correctly, 11 students (frequency) spelled 7 words cor-

rectly, 13 students (frequency) spelled 6 words correctly, 17 students (frequency) spelled 5 words correctly, 20 students (frequency) spelled 4 words correctly, 17 students (frequency) spelled 3 words correctly, 11 students (frequency) spelled 2 words correctly, and 4 students (frequency) spelled 1 word correctly. Now, the frequency distribution for these 100 middle students would look like this:

Example 1

X	f
9	2
8	5
7	11
6	13
5	17
4	20
3	17
2	11
1	4
$N = \Sigma f =$	100

The symbol "X" is used to replace the scores, and the "f" represents the abbreviation of frequency. The nice thing about grouped data into a frequency distribution is that it makes it easy to understand, and the data can be easily analyzed by the TI-30xa calculator.

Exercise 1.1

The following exercise has data from 20 middle school students who took a 10-item spelling test. The scores were the following: 10, 10, 9, 9, 9, 9, 9, 8, 8, 8, 8, 8, 8, 8, 7, 7, 7, 6, 6, 4. Construct frequency distribution, using 10, 9, 8, 7, 6, 5, and 4 as the X values.

Answer to Exercise 1.1

X	*f*
10	2
9	5
8	7
7	3
6	2
5	0
4	1

Notice that the summation of f($\sum$f) equals the number of students. More formally, the formula that finds the number of middle school students or the N (number of people) is the following:

(1.1) $$N = \sum f$$

Note: $\sum$–sigma, the Greek symbol, just means to add up all of the f values or numbers (2, 5, 7, 3, 2, 0, 1).

Proportions and Percentages

Not only can two columns (X and f) be obtained from a frequency distribution, but also a third–proportions and percentages can be displayed for the test scores. Proportions measure a fraction or part of 1, and percentages measure a fraction or part of 100. Additionally, a proportion can be defined as f/N. For example, from Exercise 1.1, 2 students out of 20 had a score of 6, so the proportion is 2/20, as a fraction, and .10 as a decimal. Proportions are sometimes referred to as *relative frequencies.* The percentage is defined as the proportion (p) multiplied by 100 (Gravetter & Wallnau, 1996). Therefore, the formula for a percentage is the following:

(1.2) $$\text{percentage} = p(100) \text{ or } f/N(100)$$

The proportion and percentage for Example 1.1 would look like the following:

X	f	$p = f/N$	$p(100)$	%
10	2	2/20 = .10	10	10%
9	5	5/20 = .25	25	25%
8	7	7/20 = .35	35	35%
7	3	3/20 = .15	15	15%
6	2	2/20 = .10	10	10%
5	0	0/20 = 0	0	0%
4	1	1/20 = .05	5	5%

A fifth column was added to show that numbers in column 4 also represent percentages.

One point that is often not mentioned in books on measurement is that proportions are also probabilities. First, probability can be defined as the frequency divided by N (number of students or cases). Alternatively, probability can be defined as the number of events (frequency) favoring X (in Example 1.1, the scores for X are 10, 9, 8, 7, 6, 5, and 4) divided by the total number of events (N). Moreover, probability is a value that ranges between zero (an event cannot happen) to one (certainty that an event will occur). In