3
EDITION

Tests & Measurement for People Who (Think They) Hate Tests & Measurement

Neil J. Salkind

PRAISE FOR THIS BOOK:

"It is extremely readable even with all the technical information. People who may have little confidence in their ability to comprehend this difficult subject matter are hooked by the conversational approach of the author."

Justina D. Pedante
Temple University

"My students say over and over again how useful this text was and that they will keep the book as a reference instead of selling it at the end of the semester!"

Janet A. Boberg
Northern Arizona University

"The real-world examples and journal article citations presented within each chapter are very useful. I like having the references and the ability to share the full article with students. I find the use of humor throughout the book to be a key component. Very serious topics (which can sometimes produce anxiety in students) are approached in a humorous manner that prevents students from becoming anxious as they start the material."

Susan L. Churchill
University of Nebraska-Lincoln

"[Salkind's] writing style and graphics are the keys to the successful delivery of some very difficult subject matter to some students with little or no prior knowledge of testing and measurement."

Brady K. LeVrier
Louisana Technical College

"The book is very straightforward and the author makes the concepts engaging and easy to understand."

Roseanne L. Flores
Hunter College of the City University of New York

"I think the main strength of this book is that it is very accessible for today's student. The most valuable feature of this book to me is the style of writing. It is appropriate for both undergraduates and graduates who lack a deeper understanding of assessment."

Edward Schultz
Midwestern State University

Tests & **Measurement** for People Who *(Think They)* Hate Tests & Measurement

Third Edition

In honor of the folks at Westwood, especially Gary D and Kristen B, and the River City Sharks; especially Larry, Mark the Shark, Millman, AK, and Kent and Annette for their kindness.

Tests & Measurement for People Who *(Think They)* Hate Tests & Measurement

Third Edition

NEIL J. SALKIND
University of Kansas

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne

FOR INFORMATION:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
3 Church Street
#10-04 Samsung Hub
Singapore 049483

Library of Congress Cataloging-in-Publication Data

Names: Salkind, Neil J., author.

Title: Tests & measurement for people who (think they) hate tests & measurement / Neil J. Salkind, The University of Kansas.

Other titles: Tests and measurement for people who (think they) hate tests and measurement

Description: Third edition. | Los Angeles : SAGE, [2017] | Includes index.

Identifiers: LCCN 2016052826 | ISBN 9781506368382 (pbk. : alk. paper)

Subjects: LCSH: Educational tests and measurements.

Classification: LCC LB3051 .S243 2017 | DDC 371.26—dc23
LC record available at https://lccn.loc.gov/2016052826

This book is printed on acid-free paper.

Acquisitions Editor: Helen Salmon
eLearning Editor: Katie Ancheta
Editorial Assistant: Chelsea Neve
Production Editor: Libby Larson
Copy Editor: Megan Granger
Typesetter: C&M Digitals (P) Ltd.
Proofreader: Scott Oney
Indexer: Will Ragsdale
Cover Designer: Candice Harman
Marketing Manager: Shari Countryman

17 18 19 20 21 10 9 8 7 6 5 4 3 2 1

This is the third edition of this book and I am particularly pleased that students and teachers have found it as useful as they have. As with *Statistics for People Who (Think They) Hate Statistics,* I receive a great deal of satisfaction helping others understand the kinds of material contained in these pages.

And much of what I say in this introduction is the same that I said in the introduction to the previous editions of the statistics book and the second edition of the tests and measurement book—take things slowly, listen in class, work hard, and you'll do fine.

Like teaching stats, teaching tests and measurement courses finds students generally anxious, but not very well informed about what's expected of them. Of course, like any worthwhile topic, learning about tests and measurement takes an investment of time and effort (and there is still the occasional monster for a teacher). But most of what they've heard (and where most of the anxiety comes from)—that tests and measurement is unbearably difficult—is just not true.

Thousands of fear-struck students have succeeded where they thought they would fail. They did it by taking one thing at a time, pacing themselves, seeing illustrations of basic principles as they are applied to real-life settings, and even having some fun along the way.

The result? A new set of tools and a more informed consumer and user of tests to evaluate all kinds of behaviors critical in all endeavors in the social and behavioral sciences from teaching to research to evaluation to diasgnosis.

After a great deal of trial and error and some successful and many unsuccessful attempts, I have learned to teach tests and measurement in a way that I (and many of my students) think is unintimidating and informative.

So, what's in store for you in these revised pages is basically what was in the two previous editions, but, this time, with a few pretty big changes; it's the information you need to understand what the field and study of basic tests and measurement are about. You'll learn the fundamental ideas about testing and tests, and how different types of tests are created and used.

There's some theory, but most of what we do in these pages focuses on the most practical issues facing people who use tests, such as what kinds of tests are available, what kind should be used and when, how tests are created and evaluated, and what test scores mean. There's a bit of math required but very little. Anxious about math? Get over it—no kidding.

The more advanced tests and measurement material is very important, but you won't find it here. Why? Because at this point in your studies, I want to offer you material at a level I think you can understand and learn with some reasonable amount of effort, while at the same time not be scared off from taking future courses.

So, if you are looking for the most recent and advanced controversies in the field, go find another good book from Sage Publications (I'll be glad to refer you to one). But, if you want to learn why and how tests and measurement can work, and then to understand the material you read in journal articles and what it means to you as a test taker and a test user, this is exactly the place.

Good luck, and let me know how I can improve this book to even better meet the needs of the beginning tests and measurement student. Send me a note at **njs@ku.edu.**

AND A (LITTLE) NOTE TO THE INSTRUCTOR

As I teach, and work on these books, two things strike me as being very important as they pertain to the teachers and others who adopt these books for use in their classes. I hope I can fairly address those here.

First, I applaud your efforts at teaching these materials. Although they may be easier for some students, most find the material very challenging. Your patience and hard work is appreciated by all, and if there is anything I can do to help, please send me a note.

Second, *Tests and Measurement for People Who (Think They) Hate Tests and Measurement* is not meant to be a dumbed-down book similar to others you may have seen. Nor is the title meant to convey anything other than the fact that many students new to the subject are actually very anxious about what's to come. This is not an academic version of a book for dummies or anything of its kind. I have made every effort to address students with the respect they deserve, to not patronize them, and to ensure that the material is approachable. How well I did in these regards is up to you, but I want to convey my very clear intent and feeling that this book contains the information needed in an introductory course, and even though there is some humor involved in my approach, nothing about the intent is anything other than serious. Thank you.

C. Deborah Laughton was the first editor on the *Statistics for People . . .* books, and Lisa Cuevas Shaw very competently took over when C. Deborah left SAGE. I owe more than words can express to them both. When Vicki Knight became editor, the professional treatment of me as an author, and the book as a product of great interest, was evident. Very fortunately for all of us, this continues with the new editor, Helen Salmon. So a great deal of thanks to those who managed this book in the past, and now especially to Helen for her excellent guidance. To those of you out there: successful books are of course about good content and good production and good marketing, but they are most about good relationships between authors and editors. I have been very fortunate.

More good fortune are the capable hands of Libby Larson who directed the production of this book and Meg Granger, who is one of best copy editors in the galaxy. They were great to work with and always provide gentle suggestions for changes that are always for the better.

Thanks also to others at SAGE, including Katie Ancheta, and Chelsea Neve and to Jason Love for providing the book's illustrations.

SAGE and I gratefully acknowledge the following reviewers for their contributions: Keith F. Donohue, North Dakota State University; Roseanne L. Flores, Hunter College of the City University of New York; Stacy Hughey Surman, University of Alabama; Thomas G Kinsey, Northcentral University; Steven Pulos, University of Northern Colorado; Edward Schultz, Midwestern State University; Cheryl Stenmark, Angelo State University; and Warren J. White, Kansas State University.

AND NOW, ABOUT THE THIRD EDITION

Any book is always a work in progress, and this latest edition of *Tests and Measurement for People Who (Think They) Hate Tests and Measurement* is no exception. The second edition was published several years ago,

and many people have told me how helpful this book is; others have told me how they would like it to change and why. In revising this book, I am trying to meet the needs of all audiences. Some things remain the same, and some have indeed changed.

When a textbook is revised, the author looks for new topics that should be covered and even old ones that have become more familiar and newly popular. In any case, there's always much more to learn and I have tried to select topics that fit.

The biggest changes in the new edition are as follows . . .

- A new chapter (Chapter 6) on item response theory, an alternative to classical test theory, that focuses more on patterns of response than just individual items. This is a whole different perspective and relatively new approach and turns out to be very interesting and applicable to may different content areas.
- The discussion on personality tests in Chapter 15 also includes extended coverage of the definition of, and uses of, neuropsychological testing.
- Inclusion of Standards for Educational and Psychological Testing which appears in Appendix D.
- Additional end-of-chapter exercises.
- A complete review of all the text with corrections and some embellishments as needed and when possible.
- Updated and expanded digital resources at study.sagepub.com/salkindtm3e.

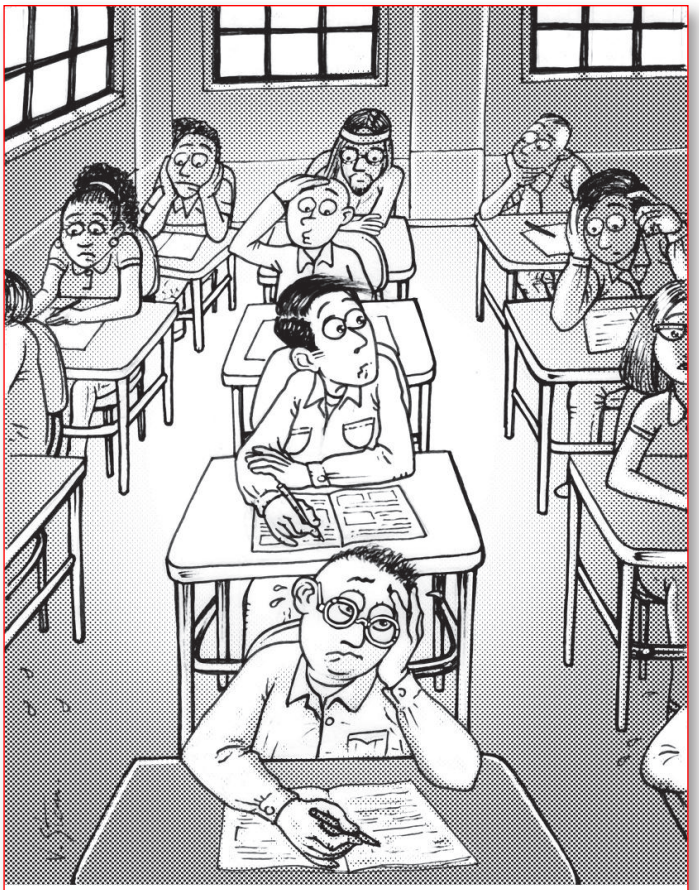
And, at the end of each chapter, you'll again find a "Real World" section that summarizes material about the chapter you just read through reviews of existing journal articles, sort of "ripped from the headlines," but not quite. These show the reader how scholarly work in the area of tests and measurement is applied to everyday concerns that we all have regarding the assessment of behavior.

Any typos and such that appear in this edition of the book are entirely my fault, and I apologize to the professors and students who are inconvenienced by their appearance. And I so appreciate any letters, calls, and e-mails pointing out these errors. You can see all these errors at **www.statisticsforpeople.com** (plus a terrific brownie recipe—no kidding), and I welcome any and all additions. We have all made every effort in this edition to correct any previous errors and hope we did a reasonably good job. Let me hear from you with suggestions, criticisms, nice notes, and so on. Good luck.

Neil J. Salkind
University of Kansas
njs@ku.edu

Neil J. Salkind received his PhD in human development from the University of Maryland, and after teaching for 35 years at the University of Kansas, he remains as Professor Emeritus in the Department of Educational Psychology and Research, where he continues to collaborate with colleagues and work with students. His early interests were in the area of children's cognitive development, and after research in the areas of cognitive style and (what was then known as) hyperactivity, he was a postdoctoral fellow at the University of North Carolina's Bush Center for Child and Family Policy. His work then changed direction to focus on child and family policy, specifically the impact of alternative forms of public support on various child and family outcomes. He has delivered more than 150 professional papers and presentations; has written more than 100 trade and textbooks; and is the author of *Statistics for People Who (Think They) Hate Statistics* (SAGE), *Theories of Human Development* (SAGE), and *Exploring Research* (Prentice Hall). He has edited several encyclopedias, including the *Encyclopedia of Human Development,* the *Encyclopedia of Measurement and Statistics,* and the recently published *Encyclopedia of Research Design*. He was editor of *Child Development Abstracts and Bibliography* for 13 years. He lives in Lawrence, Kansas, where he likes to read, swim with the River City Sharks, work as the proprietor and sole employee of big boy press, bake brownies (see www.statisticsforpeople.com for the recipe), and poke around old Volvos and old houses.

PART I



Source: Jason Love

(Sounds of thunder and lightning)

And in the beginning, there was . . . tests and measurement. Not really, but you will be surprised to learn shortly how early this measurement thing started. That's what the first part of *Tests & Measurement for People Who (Think They) Hate Tests & Measurement* is all about—a little history, an introduction to what kinds of tests there are and what they are used for, and then something about how to use this book.

You're probably new to this, and I'm sure you couldn't wait for this course to begin ☺. Well, it's here now, and believe it or not, there's a lot to learn that can be instructive and even fun—and immeasurably valuable. Let's get to it.

It's been happening to you, and you've been doing it since you were very young—being tested and taking tests.

When you were born, the doctor administered the APGAR to assess your *A*ppearance (or color), *P*ulse (or heart rate), *G*rimace (or response to stimulation), *A*ctivity (or muscle tone), and *R*espiration (or respiration). You were also screened (and it's the law in almost every state) for certain types of metabolic disorders (such as PKU or phenylketonuria)—and that may have been tests number one and two.

Then there may have been personality tests (see Chapter 15), spelling tests (see Chapter 14), statewide tests of educational progress (see Chapter 20), the ACT (American College Test) or the SAT (which actually is not an acronym—see Chapter 16 for more on this), and maybe even the GRE (Graduate Record Exam). Along the way, you might have received some career counseling using the SVIB (Strong Vocational Interest Blank) and perhaps a personality test or two such as the MMPI (Minnesota Multiphasic Personality Inventory) or the Myers-Briggs Type Inventory.

My, that's a lot of testing, and you're nowhere near done.

You've still probably got a test or two to complete once you graduate from school, perhaps as part of a job application, for additional studies, or for screening for a highly sensitive job as a secret agent.

Testing is ubiquitous in our society, and you can't pick up a copy of the *New York Times, Chicago Tribune,* or *Los Angeles Times* without finding an article about testing and some associated controversy.

The purpose of *Tests & Measurement for People Who (Think They) Hate Tests & Measurement* is to provide an overview of the many different facets of testing, including a definition of what tests and measurement is as a discipline and why it is important to study;

the design of tests; the use of tests; and some of the basic social, political, and legal issues that the process of testing involves. And when we use the word *test*, we are referring to any type of assessment tool, assessing a multitude of behaviors or outcomes.

This first part of *Tests & Measurement for People Who (Think They) Hate Tests & Measurement* will familiarize you with a basic history of testing and what the major topics are that we as teachers, nurses, social workers, psychologists, parents, and human resource managers need to understand to best negotiate our way through the maze of assessment that is a personal and professional part of our lives.

Let's start at the beginning and take a brief look at what we know about the practice of testing and how we got to where we are.

A FIVE-MINUTE HISTORY OF TESTING

First, you can follow all this history stuff by using the cool time line for what happened when, beginning at the bottom of this page and appearing throughout the chapter. Here's a summary.

Imagine this. It's about 2200 years BCE (Before the Common Era), and you're a young citizen living in a large city in China looking for work. You get up, have some breakfast, walk over to the local "testing bureau," and sit down and take a test for what we now know as a civil service position (such as a mail carrier). And at that time, you had to be proficient in such things as writing, arithmetic, horsemanship, and even archery to be considered for such a position. Must have been an interesting mail route.

Yep—testing in one form or another started that long ago, and for almost 3,000 years in China, this open (anyone could participate), competitive (only the best got the job) system proved to be the model for later systems of evaluating and placing individuals (such as the American and British civil service systems that started around 1889 and 1830, respectively).

Interestingly, this system of selection was abandoned in China around the turn of the 20th century, but we know from our own experience that the use of testing for all different purposes has grown rapidly.

Way back (around 2200 BCE)

Public officials tested in China

How Much to Take That Test? Testing is on the increase by leaps and bounds (see Chapter 20), and it's not getting any cheaper. The Brookings Institution Washington think tank estimates that $1.7 billion was spent on assessment for only the K–12 crowd—no college-level testing included. That's a ton of money, and the entire endeavor is expected to get even more expensive as the federal government moves toward expanding standardized testing to more grades in the near future.

Not much of a formal or recorded nature occurred before the middle of the 19th century, and by about the end of the 19th century, along comes our friend Charles Darwin, whom you may know from some of your other classes as the author of the *Origin of Species* (available in the first edition for only about $185,000 at the time of this writing). This book (of which only 11 copies of the first edition have survived) is a groundbreaking work that stressed the importance of what he called "descent with modification" (which we now call evolution). His thesis was that through the process of variation, certain traits and attributes are selected (that is, they survive while others die out), and these traits or attributes are passed on from generation to generation as organisms adapt.

So why are we talking about Charles Darwin and biology in a tests and measurement book? Two reasons.

First, Darwin's work led to an increased interest in and emphasis on individual differences—and that's what most tests examine. And second, Darwin's cousin (how's that for a transition?) Francis Galton was the first person to devise a set of tools for assessing individual differences in his anthropometric lab, where one could have all kinds of variables measured, such as height, weight, strength, and even how steady you can hold your hands. His motto was "Wherever you can, count." And by the way, Sherlock Holmes's motto was "Data! Data! Data!" They must have been *very* busy guys.

Once physical measurements were being made regularly, it was not long before such noted psychologists as James Cattell were working on the first "mental test." Cattell was a founder of the Psychological Corporation in the early 1920s, now known as one of the leading publishers of tests throughout the world.

1644

"If something exists, it exists in some amount. If it exists in some amount, then it is capable of being measured." Rene Descartes, Principles of Philosophy

When we get to the 20th century, testing and measurement activity really picks up. There was a huge increase in interest devoted to mental testing, which shortly became known as intelligence testing and also included the testing of cognitive abilities such as memory and comprehension. More about this in Chapter 17.

A major event in the history of testing occurred around 1905, when Alfred Binet (who was then the Minister of Public Instruction in Paris) started applying some of these new tools to the assessment of Parisian schoolchildren who were not performing as well as expected. Along with his partner, Theodore Simon, Binet used tests of intelligence in a variety of settings—and for different purposes—beyond just evaluating schoolchildren's abilities. Their work came to America in about 1916 and was extended by Lewis Terman at Stanford University, which is probably why one of the most commonly used modern intelligence tests is named the Stanford–Binet.

As always, necessity is the mother and father of invention, and come World War II, there was a huge increase in the need to test and classify accurately those thousands of (primarily) men who were to join the armed services. This occurred around World War I as well, but with nowhere near the same amount of scientific deliberation.

And as always, intense efforts at development within the government usually spill over to civilian life, and after the war (World War II, that is), hundreds of different types of tests were available for use in the civilian sector and made their way into hospitals, schools, and businesses. Indeed, we have come a long way from spelling tests.

While all these mental and ability tests were being developed, increased attention was also being paid to other dimensions of psychological functioning, such as personality development (see Chapter 15). People might be smart (or not smart), but psychologists also wanted to know how well adjusted they were and whether they were emotionally mature enough to assume certain important responsibilities. Hence, the field of personality testing (around World War I) got started in earnest and certainly is now a major component of the whole field of tests and measurement.

But our brief history of testing does not stop with intelligence or personality testing. As education became more important, so did evaluating achievement (see Chapter 14). For example, in 1937, the then-called Stanford Achievement Tests (or SATs) became required for admission to

Ivy League schools (places such as Brown, Yale, and Princeton)—with more than 2,000 high school seniors taking the exam. Another example? In 1948, the Educational Testing Service (known as ETS) opened, almost solely to emphasize the assessment of areas other than intelligence. They are the folks that bring you today's SAT, GRE, and the always popular and lovable Test of English as a Foreign Language (or TOEFL)—all taken by hundreds of thousands of students each year.

Now thousands upon thousands of high school students take standardized tests at the beginning of their senior year, and so do college seniors trying to gain admission to medical, law, and other graduate programs.

It's no wonder that services offering (and sometimes guaranteeing) success began to proliferate around 1945 with Stanley Kaplan. A very smart New Yorker (who was denied admission to medical school), he started tutoring students in the basement of his home for $0.25 per hour. His success (and it's still a hotly debated issue whether you can indeed raise people's scores through instruction) led him to create an empire of test centers (sold off for a bunch of millions to a big test company) that is still successful today.

Today, thousands and thousands of tests (and hundreds of test publishers—see Appendix B) measure everything from Advanced Placement Examination in Studio Art, which is designed to measure college-level achievements in studio arts, to the Health Problems Checklist, which is used to assess the health status and potential health problems of clients in psychotherapy settings.

And a new emphasis on the study of neuroscience has led to new evaluative efforts that explore and assess the impact of brain behavior on performance and an intense look at the role and function of testing—not without a great deal of controversy about topics such as online testing, fair testing using a common core as the basis for educational valuation, high-stakes testing, and more.

SO, WHY TESTS AND MEASUREMENT?

This question has a pretty simple answer, but simple does not mean lacking in complexity or significant implications.

No matter what profession we enter, be it teaching, social work, nursing, or any one of thousands more, we are required to make

judgments every day, every hour, and in some cases, every few minutes about our work. We do it so often that it becomes second nature. We even do it automatically.

In the most straightforward of terms, we use a test (be it formal or informal) to measure an outcome and make sense of that judgment. And because we are smart, we want to be able to communicate that information to others. So if we find that Russ got 100% on a spelling test or a 34 on his ACTs, we want everyone who looks at that score to know exactly what it means.

For example, consider the teacher who records a child's poor grade in math and sends home some remedial work that same evening; the nurse who sees a patient shivering and takes his or her temperature; or the licensed clinical social worker who recognizes client has significant difficulties concentrating and administers a test to evaluate that client's ability to stay on task and, based on the score, designs an intervention. These people all recognize a symptom of something that has to be looked into further, and they take appropriate action.

What all these professionals have in common is that in order for them to take action to help the people with whom they work, they need to first assess a particular behavior or set of behaviors. And to make that assessment, they use some kind of formal test (such as a standardized test in the case of the nurse) or informal test (such as in the teacher's case) to complete an assessment. Then, based on their training and experience, they make a decision as to what course of action to take.

For our purposes here, we are going to define a **test** as a (pick any of the following) tool, procedure, device, examination, investigation, assessment, or measure of an outcome (which is usually some kind of behavior). A test can take the form of a 50-question, multiple-choice history exam or a 30-minute interview of a parent's relationships with his or her children. It can be a set of tasks that examine how good someone is at fitting together blocks into particular designs, or whether they prefer multigrain Cheerios® to plain Cheerios®. We use tests that come in many different forms to measure many different things.

What We Test

We test many, many different things, and the thousands of tests that are available today cover a wide range of areas. Here's a quick review

of some of the content areas that tests cover. We'll go into greater detail in each of these in Part IV of *Tests & Measurement for People Who (Think They) Hate Tests & Measurement*.

We'll define these different general areas here, and in Table 1.1 you can see a summary along with some real-world examples.

Achievement tests assess an individual's level of knowledge in a particular domain. For example, your midterm in history was an achievement test.

Personality tests (covered in Chapter 15) assess an individual's unique and stable set of characteristics, traits, or attitudes. You may have taken an inventory that determined your level of introversion or extraversion.

Aptitude tests (covered in Chapter 16) measure an individual's potential to succeed in an activity requiring a particular skill or set of skills. For example, you may take an aptitude test that assesses your potential for being a successful salesperson.

Ability or intelligence tests (covered in Chapter 17) assess one's level of skill or competence in a wide variety of areas. For example, intelligence tests are viewed as measures of ability (but don't be fooled by the name of a test—there are plenty of intelligence tests that are also seen as being aptitude tests—see the following box!).

Neuropsychological tests (covered in Chapter 15) assess the functioning of the brain as it relates to everyday behaviors, including emotions and thinking.

Finally, **vocational or career tests** (covered in Chapter 18) assess an individual's interests and help classify those interests as they relate to particular jobs and careers. For example, you may have taken a vocational test that evaluates your level of interest in the culinary arts or the health care professions.

Just What Test Is That? There is always a great deal of overlap in the way people categorize particular types of tests and what they assess. For example, some people consider intelligence to be an ability (and would place it under ability tests), whereas others think of it as an achievement test because it tests one's knowledge about a particular area of information. Or aptitude tests can end up as ability tests as well as personality tests, or they can stand all on their own.

1926

The College Board publishes the Scholastic Aptitude Test

1927

Carl Spearman's notion of a general and specific factor theory of intelligence

Table 1.1 An Overview of What We Test and Some Examples of Such Tests

Type of Test	What It Measures	Some Examples
Achievement	Level of knowledge in a particular domain	Closed High School Placement Test Early School Assessment Norris Educational Achievement Tests Test of Basic Adult Education
Personality	Unique and stable set of characteristics, traits, or attitudes	Achievement/Motivation Profile Aggression Questionnaire Basic Living Skills Scale Dissociative Features Profile Inventory of Positive Thinking Traits
Aptitude	Potential to succeed	Differential aptitude tests Scholastic Aptitude Scale Aptitude Interest Category Evaluation Aptitude Test Wilson Driver Selection Test
Ability or intelligence	Skill or competence	Wechsler Intelligence Scale for Children Stanford–Binet Intelligence Test Cognitive Abilities Test General clerical ability tests School Readiness test
Performance	Basic performance of particular tasks	Achenbach System of Empirically Based Assessment Assessment in Nursery Education Functional Communication Profile The Egan Bus Puzzle Test
Vocational or career	Job-related interests	Adaptive Functioning Index Career Interest Inventory Prevocational Assessment Screen Rothwell-Miller Interest Blank Vocational Adaptation Rating Scales
Neuropsychological tests		Boston Naming Test Cognitive Symptoms Checklist d2 Test of Attention Kaplan Baycrest Neurocognitive Assessment Ruff Figural Fluency Test

Note: You can find out more about many of these tests by going to the Buros Center for Testing.

1938

Mental Measurements Yearbook first published

1939

Wechsler-Bellevue Intelligence Scale developed

So what's right? They are all right. The way we classify tests is strictly a matter of organization and convenience, and even a matter of how they are used. The definitions and examples given here reflect the current thinking about tests and measurement. Others feel differently. Welcome to the real world.

Why We Test

Now you know that there are different forms of tests and that there are many different areas of human performance and behavior that are tested regularly. But for what purpose? Here's a summary of the five main purposes (and there are surely more) for which tests can be used.

Tests are used for *selection*. Not everyone can be a jet pilot, so only those men or women who score at a certain level of performance on physical and psychological assessments will be selected for training.

Tests are used for *placement*. Upon entering college, not everyone should be in the most advanced math class or in the most basic. A placement test will determine where the individual belongs.

Tests are used for *diagnosis*. An adult might seek out psychological counseling, and the psychologist may administer a test or group of tests that helps diagnose any one of many different mental disorders. Diagnostic tests are also used to identify individual strengths and weaknesses.

Tests are used for *hypothesis testing*. A hypothesis is simply an "if . . . then" statement. For example, if children get extra reading help throughout the week, then they will score better on a reading test of comprehension than will children who do not get extra help. One important part of testing this question is using a test that measures reading comprehension accurately.

Finally, tests are used to *classify*. Want to know what profession might suit you best? One of several different tests can provide you with an idea of your aptitude (or future potential) for a career in the culinary arts, auto mechanics, medicine, or child care.

What Are Tests Used For? Tests are used widely for a variety of purposes, among them selection, placement, diagnosis, hypothesis testing, and classification.

SOME IMPORTANT REMINDERS

You'll learn many different things throughout *Tests & Measurement for People Who (Think They) Hate Tests & Measurement* (at least we sure hope you will). And with any vibrant and changing discipline, there are always discussions both pro and con about different aspects of the subject. But there are some constants as well, as presented below.

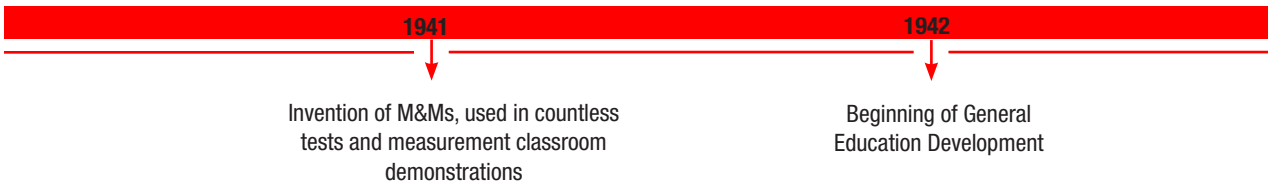
1. *Some behaviors can be observed more closely and more precisely than others.* It's pretty easy to measure one's ability to add single digits (such as 6 + 5 = ?), but to understand *how* one solves (not *if* one can solve) a quadratic equation is a different story. The less obvious behaviors take a bit more ingenuity to measure, but that's part of the challenge (and delight) of doing this.

2. *Our understanding of behavior is only as good as the tools we use to measure it.* There are all kinds of ways we try to measure outcomes, and sometimes we use the very best instruments available—and at other times, we may just use what's convenient. The development and use of the best tools takes more time, work, and money, but it gives us more accurate and reliable results. Anything short of the best forces us to compromise, and what you see may, indeed, not be what you get.

No matter how interesting your theory or approach to a problem, what you learn about behavior is only as accurate and worthwhile as the integrity and usefulness of the tools you use to measure that behavior.

3. *Tests and measurement tools can take many different forms.* A test can be paper and pencil, self-report, observation, or performance and often gives us very similar information on some outcome in which we are interested. And in some cases, tests are restricted by what they are measuring. For example, most achievement tests are paper and pencil, and most tests that look at performance of motor skills are just that, performance. The lesson here is to select the form of test that best fits the question you are asking.

4. *The results of any test should always be interpreted within the context in which they were collected.* In many communities, selected



junior high students take a practice Scholastic Assessment Test. Although some of these students do very, very well, others perform far below what you would expect a high school junior or senior to do; perhaps these younger children simply have not yet had the course work. To interpret the results of the younger children using the same metric and scoring standards as for the older children would surely not do either group any justice. The point is to keep test scores in perspective—and of course, to understand them within the initial purpose for the testing.

5. *Test results often can be misused.* It doesn't take a rocket scientist to know that there have been significant controversies over how tests are used. You'll learn more about this in Part V of *Tests & Measurement for People Who (Think They) Hate Tests & Measurement,* but many of you know how non-English-speaking immigrants who tried to get sanctuary in the United States were turned away in the 1930s based on their test scores. To use tests fairly and effectively, you need to know the purpose of the test, the quality of the test, how it is administered and used, and how the results are interpreted. We'll do all that in *Tests & Measurement for People Who (Think They) Hate Tests & Measurement.*

6. *Many tests, especially achievement tests, have as their goal distinguishing between those who know the material and those who do not.* We want the biology student to understand evolution and the sixth grader to know something about American and world history.

HOW TESTS ARE CREATED

We can suggest several books that are all about the theory and mechanics of test construction, and this is not one of them. But rather than reading 400 pages about this important topic, we're going to offer a summary of how, in general, a test is designed and which steps are part of the process.

The entirety of the process shown in Table 1.2 is linear; that is, step 2 always follows step 1, but within each step there is some evaluation of whether it is time to move on to the next step. Let's take a look.

Table 1.2 A Broad Description of the Steps in the Development of a Test

Step 1

Idea, trait, or characteristic to be tested

↓

Step 2

Best method (paper and pencil, performance, survey, interview)

to assess the focus of the test

↓

Step 3

The development of items

↓

Step 4

The pilot testing of items

↓

Step 5

The evaluation and revision of items

that are not satisfactory

↓

Step 6

The pilot testing of items

↓

Step 7

Final determination of items

↓

Step 8

Development and revision of

administration instructions

↓

Step 9

Establishment of reliability and validity

↓

Step 10

Revision of items in light of validity and reliability studies and

creation of norms if necessary

↓

Step 11

Development of norms

↓

Step 12

Completion of the test manual

So What's New?

Up to now, the development of most tests falls within a classical test theory (or CTT) model. The CTT model (and most of this book discusses the various aspects of that model) primarily looks to increase the accuracy of predicting a test taker's true score or the actual value of a trait, characteristic, level of knowledge, or any other domain. As you will learn later, true scores are theoretical in nature, because it is impossible to rule out all sources of error in test taking (such as bad instructions, ill-prepared test takers, etc.) and get a totally, 100% accurate true score. All these sources of error contribute to an individual's final score.

At least one alternative to CTT is item response theory (or IRT), which places the emphasis not on the individual's performance and the accompanying sources of error but on the items and how item difficulty is not a constant and can change.

We'll distinguish between CTT and IRT (as well as some other new approaches) in Chapter 6. All you need to know for now is that, as in almost all disciplines, new ideas and techniques are always being developed, almost always interesting, and surely always ripe for discussion and friendly differences among experts, colleagues, and students as to what's best.

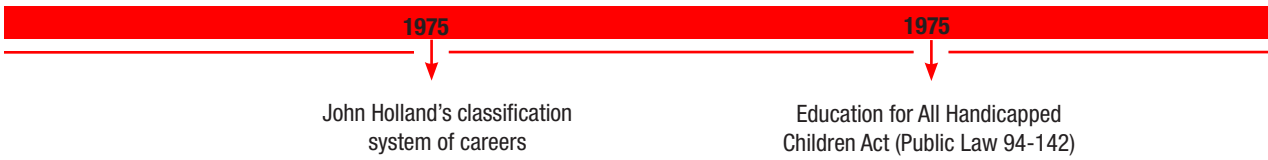
WHAT AM I DOING IN A TESTS AND MEASUREMENT CLASS?

There are probably many reasons why you find yourself using this book. You might be enrolled in an introductory tests and measurement class. You might be reviewing for your comprehensive exams. Or you might even be reading this on summer vacation (horrors!) in preparation and review for a more advanced class.

In any case, you're a tests and measurement student whether you have to take a final exam at the end of a formal course or whether you're just in it of your own accord. But there are plenty of good reasons to be studying this material—some fun, some serious, and some both.

Here's a list of some of the things my students hear at the beginning of our introductory tests and measurement course.

1. Tests and Measurement 101 or Introduction to Testing or whatever it's called at your school looks great listed on your transcript. Kidding aside, this may be a required course for you to complete your major. But even if it is not, having these skills is definitely a big plus when it comes time to apply for a job or for further schooling. And with more advanced courses, your résumé will be even more impressive.
2. If this is not a required course, taking a basic tests and measurement course sets you apart from those who do not. It shows that you are willing to undertake a course that is above average in regard to difficulty and commitment.
3. Basic information about tests and measurement is an intellectual challenge of a kind that you might not be used to. A good deal of thinking is required, as well as some integration of ideas



and application. The bottom line is that all this activity adds up to what can be an invigorating intellectual experience, because you learn about a whole new area or discipline.

4. There's no question that having some background in tests and measurement makes you a better student in the social, behavioral, and health sciences. Once you have mastered this material, you will have a better understanding of what you read in journals and also what your professors and colleagues may be discussing and doing in and out of class. You will be amazed the first time you say to yourself, "Wow, I actually understand what they're talking about." And it will happen over and over again, because you will have the basic tools necessary to understand exactly how scientists reach the conclusions they do.
5. If you plan to pursue a graduate degree in education, anthropology, economics, nursing, medicine, sociology, or any one of many social, behavioral, and health sciences fields, this course will give you the foundation you need to move further.
6. Finally, you can brag that you completed a course that everyone thinks is the equivalent of building and running a nuclear reactor.

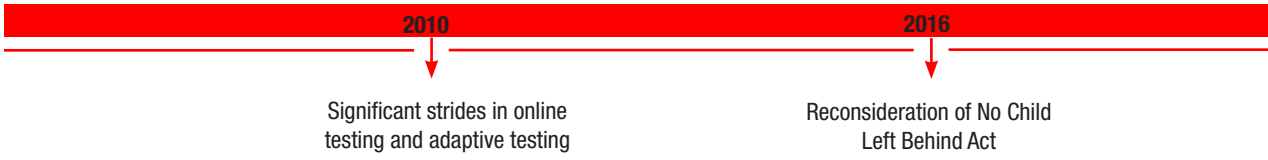
TEN WAYS TO USE THIS BOOK (AND LEARN ABOUT TESTS AND MEASUREMENT AT THE SAME TIME!)

Yep. Just what the world needs—another tests and measurement book. But this one is different. It's directed at the student, is not condescending, is informative, and is as simple as possible in its presentation. It assumes only the most basic information at the start, and if you don't have that, you can go to Appendix A and get it.

However, there has always been a general aura surrounding the study of tests and measurement that it's a difficult subject to master. And I don't say otherwise, because parts of it are challenging. On the other hand, millions and millions of students have mastered this topic, and you can, too. Here are a few hints to close this introductory chapter before we move on to our first topic.

- *You're not dumb.* That's true. If you were, you would not have gotten this far in school. So treat tests and measurement like any other new course. Attend the lectures, study the material, and do the exercises in the book and from class, and you'll do fine. Rocket scientists know how to use this stuff, but you don't have to be a rocket scientist to succeed.
- *How do you know tests and measurement is hard?* Is this topic difficult? Yes and no. If you listen to friends who have taken the course and didn't work hard and didn't do well, they'll surely volunteer to tell you how hard it was and how much of a disaster it made of their entire semester, if not their lives. And let's not forget—we always tend to hear from complainers. So I suggest that you start this course with the attitude that you'll wait and see how it is and judge the experience for yourself. Better yet, talk to several people who have had the class and get a good general idea of what they think. Just don't base your opinion on one spoilsport's experience.
- *Form a study group.* This is one of the most basic ways to ensure some success in this course. Early in the semester, arrange to study with friends. If you don't have any who are in the same class as you, then make some new ones or offer to study with someone who looks to be as happy about being there as you are. Studying with others allows you to help them if you know the material better, or to benefit from others who know the material better than you do. Set a specific time each week to get together for an hour and go over the exercises at the end of the chapter or ask questions of one another. Take as much time as you need. Find a coffee shop and go there with your study buddy. Studying with others is an invaluable way to help you understand and master the material in this course.

Stay on Task and Take One Thing at a Time. Material about testing and measurement can be tough to understand, especially if you have never heard any of these terms before or thought about any of these ideas. Follow the guidelines mentioned here and talk with your teacher as soon as you find yourself not understanding something or falling behind.



Ask your teacher questions, and then ask a friend. If you do not understand what you are being taught in class, ask your professor to clarify it. Have no doubt—if you don't understand the material, then you can be sure that others do not as well. More often than not, instructors welcome questions. And especially because you've read the material before class, your questions should be well informed and help everyone in class better understand the material.

Do the exercises at the end of a chapter. The exercises are based on the material and the examples in the chapter they follow. They are there to help you apply the concepts that were taught in the chapter and build your confidence at the same time. How do the exercises do that? An explanation for how each exercise is solved accompanies the problem. If you can answer these end-of-chapter exercises, then you are well on your way to mastering the content of the chapter.

Practice, practice, practice. Yes, it's a very old joke:

Q. How do you get to Carnegie Hall?
A. Practice, practice, practice.

Well, it's no different with basic statistics. You have to use what you learn and use it frequently to master the different ideas and techniques. This means doing the exercises in the back of the chapter as well as taking advantage of any other opportunities you have to understand what you have learned.

Look for applications to make it more real. In your other classes, you probably have occasion to read journal articles, talk about the results of research, and generally discuss the importance of the scientific method in your own area of study. These are all opportunities to look and see how your study of tests and measurement can help you better understand the topics under class discussion as well as the area of beginning statistics. The more you apply these new ideas, the better and more full your understanding will be.

Browse. Read the assigned chapter first, then go back and read it with more intention. Take a nice leisurely tour of *Tests & Measurement for People Who (Think They) Hate Tests & Measurement* to see what's contained in the various chapters. Don't rush yourself. It's always good to know what topics lie ahead, as well as to familiarize yourself with the content that will be covered in your current statistics class.

Have fun. This indeed might seem like a strange thing for you to read, but it all boils down to your mastering this topic rather than letting the course and its demands master you. Set up a study schedule and

follow it, ask questions in class, and consider this intellectual exercise to be one of growth. Mastering new material is always exciting and satisfying; it's part of the human spirit. You can experience the same satisfaction here. Just keep your eye on the ball and make the necessary commitment to stay current with the assignments and work hard.

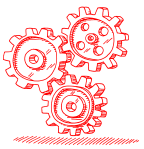
Finally, be easy on yourself. This is not material that any introductory student masters in a matter of hours or days. It takes some thinking and some hard work, and your expectations should be realistic. Expect to succeed in the course, and you will.

About Those Icons

An icon is a symbol. Throughout *Tests & Measurement for People Who (Think They) Hate Tests & Measurement*, you'll see a variety of different icons.

Here's what each one is and what each represents:

This icon represents information that goes beyond the regular text. It might be necessary to elaborate on a particular point, and that can be done more easily outside the flow of the usual material.



In Tech Talk, I discuss some more technical ideas and tips to inform you about what's beyond the scope of this course. You might find these interesting and useful.

Every now and then, but not often, you'll find steps like the ones you see here. This indicates that there is a set of steps coming up that will direct you through a particular process. These steps have been tested and approved by whatever federal agency approves these things.

That finger with the bow is a cute icon, but its primary purpose is to help reinforce important points about the topic you just read about. Try to emphasize these points in your studying, because they are usually central to the topic.

The Famous Difficulty Index

For want of a better way to give you some upfront idea about the difficulty of the chapter you are about to read, we have developed a

highly secret difficulty index using smileys. This lets you know what to expect as you begin reading.

How Hard Is This Chapter?	*Look at Mr. Smiley!*
Very hard	☺
Hard	☺☺
Not too hard, but not easy either	☺☺☺
Easy	☺☺☺☺
Very easy	☺☺☺☺☺

GLOSSARY

Bolded terms in the text are included in the glossary at the back of the book.

SUMMARY

Now you have some idea about what a test is and what it does, what areas of human behavior are tested, and even the names of a few tests you can throw around at tonight's dinner table. But most of all, we introduced you to a few of the major content areas we will be focusing on throughout *Tests & Measurement for People Who (Think They) Hate Tests & Measurement.*

TIME TO PRACTICE

1. What are some of your memories of being tested? Be sure to include (if you can) the nature of the test itself, the settings under which the test took place, how prepared or unprepared you felt, and your response upon finding out your score.
2. Go to the library (not to the Internet) and identify five journal articles in your area of specialization, such as teaching math or nursing or social work. Now create a chart like this for each set of five.

Journal Name	Title of Article	What Was Tested	What Test Was Used to Test It?

   a. Were most of the tests used developed commercially, or were they developed just for this study?

   b. Which test do you think is the most interesting, and why?

   c. Which test do you think got the closest to the behavior that the authors wanted to measure?

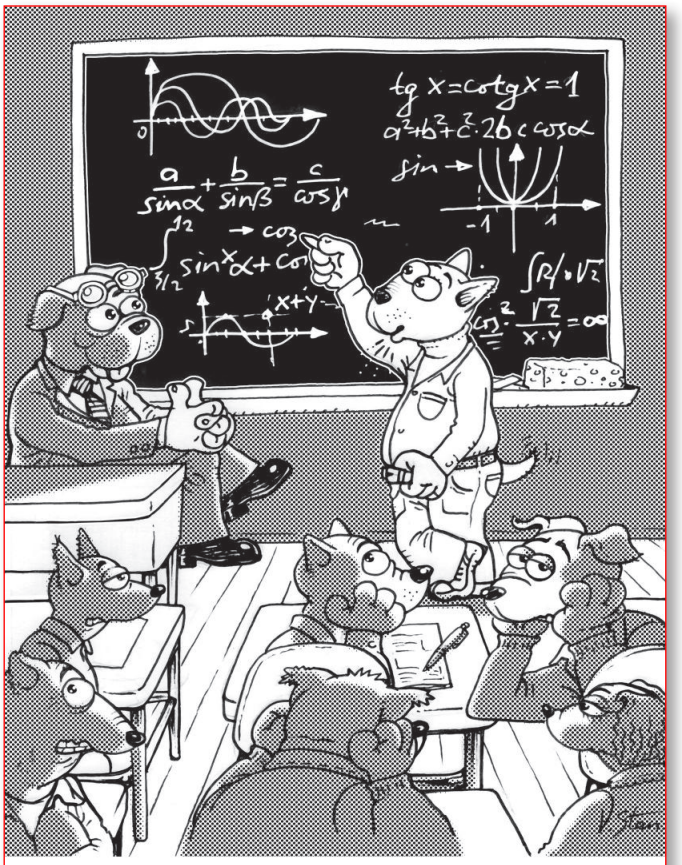
3. Ask your parent, child, professor, colleague, or classmate what he or she believes are the most important reasons for testing and what types of tests he or she can identify.

4. One of the things we did in this opening chapter was identify five different purposes of tests (see page 11). Think of at least two other ways that tests might be used, and give a real-world example of each.

5. Interview someone who uses tests in his or her work, as either an assessment or a research tool, and try to get an idea of the importance he or she places on being knowledgeable about testing and what role it plays in his or her research and everyday professional career. Is he or she convinced that tests assess behavior fairly? Does he or she use alternatives to traditional testing? Does he or she find the results of tests useful for helping students?

6. Extra credit and extra imagination: Use your favorite search engine and search on five different topics related to testing in general, such as fairness in testing, use of computerized testing, how tests are developed, and so on. Use your imagination and search as broadly as possible. Summarize the results of these searches and propose some directions you think testing might be taking in future activities.

PART II



Cronbach's Alpha Dog.

Source: Jason Love

In this part of *Tests & Measurement for People Who (Think They) Hate Tests & Measurement,* we discuss some of the most important and fundamental ideas that provide the foundation for developing and using tests.

In Chapter 2, we review levels of measurement—what they are and how they are used. Here, you'll learn how different levels of measurement coincide with different amounts of information that a particular measure conveys.

Our study of reliability, in Chapter 3, takes us on a tour of the idea of how consistent a test is. This chapter brings us information on the conceptual nature of what reliability is, the many different types of reliability, and how they are used, when each is appropriate to use, and how each is established in practice.

Probably the most important construct in the study of tests and measurement is validity—reliability's first cousin—or whether a test does what it is supposed to do. What validity is, the different types of validity, and how validity is established are all covered in Chapter 4.

Chapter 5 provides us with a review of test scores and how to best understand them, and Chapter 6—new to this edition—introduces us to the idea of item response theory, an alternative to the long-established classical test theory. This "new" stuff is pretty cool.

How things are measured is very important to our study of tests and measurement. And how *precisely* they are measured is just as important. In one study, researchers from the University of Kansas examined how the very interesting concepts of self-determination and self-concept have an effect on academic achievement for adolescents with learning disabilities. Using trusted and field-tested assessment tools, they found significant relationships among the three variables of self-determination, self-concept, and academic achievement, with self-determination being a useful predictor of academic achievement for these students.

Want to know more? Take a look at Zheng, C., Erickson, A., Kingston, N., & Noonan, P. (2014). The relationship among self-determination, self-concept, and academic achievement for students with learning disabilities. *Journal of Learning Disabilities, 47*, 462–474.

FIRST THINGS FIRST

Before we start talking about levels of measurement, let's spend a moment defining a few important terms—specifically, what a variable is and what the term *measurement* means.

A **variable** is anything (such as a test score) that can take on more than one value. For example, a score on the SAT can take on more than one value (such as 750 or 420), as can what category of color your hair falls into (such as red or black). Age is another variable (someone can be 2 months old or 87 years old), as is favorite flavor of ice cream (such as rocky road or mint chocolate chip). Notice that the labels we apply to outcomes can be quantitative (such as 87 years) or qualitative (such as rocky road or mint chocolate chip). Good, that's out of the way.

Now, the term **measurement** means the assignment of labels to (you guessed it) a variable or an outcome. So when we apply the label "black" to a particular outcome, we are measuring that outcome. We can measure the number of windows in a house, the color of a car, the score on a test of memory, and how fast someone can run 100 yards. In every case, we are assigning a label to an outcome. Sometimes that label is precise (such as 10.7 seconds for the time it takes to run 100 yards), and sometimes it is less precise (such as "like" for how someone feels about a presidential candidate).

There is a great deal of discussion (and of course controversy) over what some levels mean. For example, the sex of a child at birth is designated (measured) as male or female. But the gender label of that individual is often the result of a social construct (such as male, female, or LGBTQ)—interesting, provocative for the measurement people, and challenging.

As the world turns, about 70 years ago, in 1946, S. S. Stevens, the famous experimental psychologist (not the steamship), started to wonder about how different types of variables are measured and whether the level of precision at which those variables are measured might be classified so they can be more easily understood. Even better, he asked the question of whether a system (what he called a set of rules) could be developed so the outcomes that result from the measurement process can be classified based on the *characteristics* of the variable itself, rather than the actual value or label of the variable.

And coming in at about 10 lbs., 11 ounces—the idea of levels of measurement was born.

How to Measure Variables? Variables can be measured in different ways, and the way a variable is being measured determines the level of measurement being used. For example, we can measure the height of an individual in several different ways. If we say that Group 1 is taller than Group 2, then we have chosen to measure this variable at a level that distinguishes one group from another only in rank or magnitude. On the other hand, if we chose to measure this variable at a level that distinguishes groups by a certain number of inches, that is much more precise and much more informative.

THE FOUR HORSEMEN (OR LEVELS) OF MEASUREMENT

A **level of measurement** represents how much information is being provided by the outcome measure. There are four levels of measurement—nominal, ordinal, interval, and ratio—and here's more about each.

The Nominal Level of Measurement

The **nominal level of measurement** describes a measurement system where there are differences in quality rather than quantity. These are variables that are *categorical* or *discrete* in nature. Here, outcome scores can be placed into one (and only one) category. These "labels" are qualitative in nature, and scores can be placed in one and only one category, which is why such scores are mutually exclusive. You can't be in one spelling group (the Guppies) and another spelling group (the Sharks) at the same time.

For example, Max is in the Guppies and Russ is in the Sharks. Knowing that much information tells us only that Max and Russ are in two different spelling groups—not how the groups differ, or who is better, or how many words Max or Russ can spell—just that they are in two different groups.

It's called the nominal level of measurement (after the word *nomin* in Latin, meaning name) because the only distinction we can make is that variables differ in the category in which they are placed. Measuring a variable such as spelling by group assignment (and nothing more) is similar to distinctions between Republicans and Democrats; white, black, yellow, and red folks; Chevy and Volvo drivers; and shoppers at Hy-Vee and Dillons. They differ from one another by the nature of the group to which they belong.

Want to know more about these differences? Well, you can't just by knowing what group they are in, and that's a significant limitation to the nominal level of measurement. If you want more information, then you have to dig more or define (and measure) the variable in a more precise way, which we will do shortly.

An example of a study that uses a nominal-level variable is one conducted by Rik Verhaeghe and his colleagues, detailed in a 2003 article that appeared in *Stress and Health: Journal of the International Society for the Investigation of Stress*. They examined job stress among middle-aged health care workers and its relation to absences due to sickness. One assessment that had to be made was based on the assignment of people to one of two groups—nurses or non-nurses—and that variable is being measured at the nominal level. As you can see, a participant can be in only one group at a time (they are mutually exclusive), and at this level of measurement, you can assign only a label (nurse or non-nurse).

Want to know more? Take a look at Verhaeghe, R., Mak, R., VanMaele, G., Kornitzer, M., & De Backer, G. (2003). Job stress among middle-aged health care workers and its relation to sickness absence. *Stress and Health: Journal of the International Society for the Investigation of Stress, 19*(5), 265–274.

Nominally Nominal. There's always a ton of discussion about measurement levels and their utility, starting with the definition of variables and how those variables are measured. And in some cases, there's doubt that the nominal level of measurement is a level of measurement at all rather than only a qualitative description of some outcome. As always, it's your decision to place nominal-level variables where you think they belong.

The categories in which measures can be placed on the nominal scale are always mutually exclusive. You can't be in the red preschool room and the blue preschool room at the same time.

Nominal-level measures are always qualitative (the values have no inherent meaning). Being in the red room is neither here nor there, as to being in the blue room. You're a preschooler in either room, and that's it; the room assignment says nothing about anything other than that—just which room you are in.

The Ordinal Level of Measurement

Next, we have the **ordinal level of measurement**, which describes how variables can be ordered along some type of continuum. (Get it? Ordinal, as in ordering a set of things.) So outcomes are placed in categories (like the nominal scale), but they have an order or rank to them as well, like stronger and weaker, taller and shorter, faster and slower, and so on.

For example, let's take Max and Russ again. As it turns out, Max is a better speller than Russ. Right there is the one and only necessary criterion for a measure to be at the ordinal level of measurement. It's the "better than" or "worse than" thing—some expression of the relationship between categories.

However, from better or worse, we cannot tell anything about how good a speller either Max or Russ is, because ordinal levels of measurement do not include this information. Max might get only 3 words out of 10 correct, whereas Russ might get only 2. That makes Max better but not very good, right?

But you can say that Max is a better speller than Russ and is better than Sophie (another classmate), and Sophie is a better speller than Sue—all relative statements.

Our real-world example is a study by Kathe Burkhardt and her colleagues that appeared in the journal *Behavior Change* in 2003. They examined common childhood fears in 9- to 13-year-old South African children, and one of the ways they assessed fears was

by having children rank them. In fact, the researchers found that the children's rankings of fears differed from rankings derived using a scale that attached an actual value to the fear.

Want to know more? Take a look at Burkhardt, K., Loxton, H., & Muris, P. (2003). Fears and fearfulness in South African children. *Behaviour Change, 20*(2), 94–102.

The Interval Level of Measurement

That's two levels of measurement down and two to go.

The **interval level of measurement** gives us a nice jump in the amount of information we obtain from a new level of measurement. You already know that we can assign names (nominal level) and rank (ordinal level), but it is with the interval level of measurement that we can assign a value to an outcome that is based on some underlying continuum that has equal intervals. And if there is an underlying continuum, then we can make very definite statements about someone's position (his or her score) along that continuum and his or her position relative to another person's position, including statements about such things as differences. Wow, that's a lot more complex than the earlier two levels of measurement and provides a lot more information as well.

For example, not only do we know that Max and Russ fall into two different categories of spellers (nominal) and that Max is better than Russ (interval), but we can now know *how much* better Max actually is. In fact, Max got 8 out of 10 correct, and Russ got 4 out of 10 correct. Because one of the assumptions of this level of measurement is that it is based on a scale that has equally appearing intervals (one correct, two correct, three correct, etc.), we can say that Max got four more correct than Russ. Or if Max got six correct and Russ got five correct, then Max would have gotten one more correct than Russ.

Variables, Their Measurement, and Their Interpretation. Although an interval-level scale provides much more information than an ordinal- or nominal-level scale, you have to be careful in how you interpret these scores. For example, scoring 50% higher on a history test does not mean that score represents 50% more knowledge (unless the test is a perfect, perfect, perfect representative of all the questions that could be asked). Rather, it means only that 50% more of the questions were answered correctly. We can conclude that the more questions correct, the better one is in history, but don't carry it too far and overgeneralize from a test score to an entire area of knowledge or a construct such as intelligence.

What's the big advantage of the interval level of measurement over the ordinal and nominal besides increased precision? In one word, *information*—there's much more of it when we know what a score actually is and what it means. Remember, Max could be ranked #1 in his class but get only 50% of the words correct on Friday's test. On the other hand, knowing what his exact score is relative to some type of underlying continuum provides us with an abundance of information when it comes time to make a judgment about his performance.

Jennifer Hill and her colleagues used an interval level of measurement when they used the Wechsler Intelligence Scale for Children to determine the effects of high participation in an infant early intervention program that targeted low-birth-weight premature infants. They found that infants who participated had higher scores on the Wechsler when they were 8 years old than did those who did not participate. Scores on the Wechsler are based on an underlying continuum such that a score of 100 represents two points less than a score of 102.

Want to know more? Take a look at Hill, J., Brooks Gunn, J., & Waldfogel, J. (2003). Sustained effects of high participation in an early intervention for low-birth-weight premature infants. *Developmental Psychology, 39*(4), 730–744.

The Ratio Level of Measurement

This is by far the most interesting measurement, yet the one that is least likely to be seen in the social or behavioral sciences. Why? Because the **ratio level of measurement** is characteristic of all the other scales we have already talked about, but it also includes a very important assumption—an *absolute zero* corresponding to an absence of the trait or characteristic being measured. Physical measurements such as amount of rainfall, weight, and height fall under the ratio level of measurement.

The scale and its use become interesting when we begin to look at nonphysical attributes or behaviors. For example, it is possible to receive a score of zero on a spelling test, right? You just need not spell any words correctly. But here's the big question: Does getting such a score indicate that one has no spelling ability? Of course not. It means only that on this test, no words were spelled correctly.

That's the challenge, then. Is there any trait or characteristic in the behavioral or social sciences that an individual can have a complete absence of? If there is not, then a ratio level of measurement is impossible. In fact, this is one reason why, when this category of tests and measurement is taught, the interval and ratio levels often are combined into one. We're not doing that here, because we think they are important enough to keep separated.

Now, in the physical and biological sciences, it's not as much of an issue or challenge. Consider temperature. Absolute zero is defined as –459 degrees Fahrenheit and –273 degrees Celsius (for all you metric fans)

and indicates no molecular activity. How about finger tapping as a measure of responsiveness? It is entirely possible to have no or zero finger taps. Both cases truly have a true zero. But even if someone scores zero on an intelligence test (perhaps one taken in Russian), does that mean he or she has no intelligence? Or if someone receives a zero on that spelling test, does that mean he or she cannot spell? Of course not.

So for us social and behavioral scientists (like most of you), we will rarely (if ever) see a ratio-level scale in the journal articles we review and read. The scale of measurement simply depends on how the variable is being defined and measured.

A SUMMARY: HOW LEVELS OF MEASUREMENT DIFFER

We just discussed four different levels of measurement and what some of their characteristics are. You also know by now that a more precise level of measurement has all the characteristics of an earlier level and provides more information as well.

In Table 2.1, you can see a summary table that addresses the following questions:

1. Are you measuring most of the available information?
2. Can you assign a name to the variable being measured?
3. Can you assign an order to the variable being measured?
4. Can you assign an underlying quantitative scale to the variable being measured?

Remember, the more precise your level of measurement, the more information is conveyed.

Table 2.1 A Summary of Levels of Measurement and the Characteristics That Define Them

Level of Measurement	Are You Measuring Most of the Available Information?	Can You Assign Names to the Variable Being Measured?	Can You Assign an Order to the Variable Being Measured?	Can You Assign an Underlying Quantitative Scale to the Variable Being Measured?	Can You Assign an Absolute Zero to the Variable Being Measured?
Ratio	Most	☺	☺	☺	☺
Interval	More	☺	☺	☺	☹
Ordinal	Less	☺	☺	☹	☹
Nominal	Least	☺	☹	☹	☹

This table shows us that the ratio level of measurement allows us to answer yes (☺) to these four questions, whereas the nominal level of measurement allows us to answer yes to only one.

OKAY, SO WHAT'S THE LESSON HERE?

The lesson here is that, when you can, try to select a technique for measuring a variable that allows you to use the highest level of measurement possible (most often the interval level). We want to access the most information available with the most precision while understanding that as the scale of measurement changes in precision, the way the variables are being measured will probably change in level of complexity as well, as you can see in Figure 2.1. As variables are measured in more sophisticated ways and become more complex in their definition and nature, they lend themselves better to higher **scales of measurement** (such as interval or ratio) than do variables that are less complex.

For example, as we discussed in the earlier example, one can define biological sex in newborns as male or female, but what if one defines it as the proportion of testosterone to estrogen (two common hormones in humans). Well, the "higher" level of definition demands more precise measurement tools to accompany the more precise way of measuring.

For another example, when testing the effectiveness of strength training in senior citizens, don't classify them as weak or strong after

Figure 2.1 Complexity and Precision and Scales of Measurement