Standards for Tests and Measurements in Physical Therapy Practice
Jules M Rothstein, Suzann K Campbell, John L Echternach, Alan M Jette, Harry G Knecht, Steven J Rose and on behalf of the Task Force on Standards for Measurement in Physical Therapy

The online version of this article, along with updated information and services, can be found online at: http://ptjournal.apta.org/content/71/8/589

Collections  This article, along with others on similar topics, appears in the following collection(s):
- Policies, Positions, and Standards
- Professional Issues
- Tests and Measurements

e-Letters  To submit an e-Letter on this article, click here or click on "Submit a response" in the right-hand menu under "Responses" in the online version of this article.

E-mail alerts  Sign up here to receive free e-mail alerts
Standards for Tests and Measurements in Physical Therapy Practice

Task Force Members

Members of the American Physical Therapy Association's Task Force on Standards for Measurement in Physical Therapy:

Jules M Rothstein, PhD, PT, is Professor and Head, Department of Physical Therapy (M/C 898), University of Illinois at Chicago, and Chief of Physical Therapy Services, University of Illinois Hospital, 1919 W Taylor St, Chicago, IL 60612 (USA).

Suzann K Campbell, PhD, PT, FAPTA, is Professor, Department of Physical Therapy, University of Illinois at Chicago.

John L Eckernach, EdD, PT, is Chairman, School of Community Health Professions and Physical Therapy, Old Dominion University, Norfolk, VA 23539-0288.

Alan M Jette, PhD, PT, is Senior Research Scientist, New England Research Institute Inc, 9 Galen St, Watertown, MA 02172.

Harry G Knecht, EdD, PT, is Director of Research, Orthosport Inc, 7736 W Madison St, Forest Park, IL 60130, and Associate Professor (Emeritus), Department of Physical Therapy, University of Illinois at Chicago.

Steven J Rose, PhD, PT, FAPTA, was Associate Director for Research, Division of Physical Therapy, University of Miami School of Medicine, 5915 Ponce de Leon Blvd, Coral Gables, FL 33146, at the time the Task Force on Standards for Measurement in Physical Therapy convened. Dr Rose died on April 4, 1989.

Table of Contents

Preface..................................................590
Dedication...........................................591
Introduction.........................................592
Glossary of Terms Used in the Standards .................595
Standards for Primary Test Purveyors..................598
Standards for Secondary Test Purveyors.................604
Standards for Tertiary Test Purveyors..................609
Standards for Test Users.............................613
Standards for Ensuring Integrity in Measurement Research........620
Index...................................................622

A primer on measurement, produced to complement the Standards for Tests and Measurements in Physical Therapy Practice, will be published by the American Physical Therapy Association in late 1991.


Preface

The *Standards for Tests and Measurements in Physical Therapy Practice* is a cohesive and well-organized document, complete with operational definitions and primer. During a first reading, one might readily conclude that the entire work is an academic treatise, esoteric in nature and dedicated to a select handful of clinicians with the intellectual capacity and interest to glean meaning from its organization. These Standards, however, represent far more than can be processed by a cursory glance or a preoccupied mind.

As defined by the Board of Directors of the American Physical Therapy Association (November 1984), a standard is an approved, binding, general statement of requirement used to judge quality of action or activity. As such, standards are accessible to the lay and health professional public for their scrutiny and criticism. To mature from concept to finality, therefore, implies a honing process of utmost delicateness and comprehension, for to expose our standards of measurements and tests to those who may judge us is to reveal an identity perhaps unknown to or misunderstood by such parties. At a time in which physical therapists seek greater autonomy in clinical decision making, standards indeed do become the palettes from which our destiny is stroked.

With this perspective in mind, a task force of the Committee on Research was created in 1987 to develop standards for tests and measures used commonly in physical therapy practice. This task force, coordinated by Jules Rothstein, included Suzann Campbell, John Echternach, Alan Jette, Harry Knecht, and the late Steven Rose. The group sought to produce standards that addressed logical requirements for measurement, reproduction of test results, and interpretation and use of such results. All materials were reviewed by representatives of Sections and Specialty Councils; physical therapists primarily in education, research, or practice environments; and external experts. Collectively, these professionals recognized that we are at risk in any working environment unless our tests and measurements are creditable and specifically identifiable with our clinical activities. Standards were needed to improve the quality of our practice, lend a unifying perspective to the instruction of measurements to our students, and enhance the rigor of our research activities. The standards necessary to meet these requirements had to be unique to physical therapy and easily associated with our professional skills by practitioners, faculty, and students; medical groups who have vested interests in the quality of physical rehabilitation services; and third-party payers who assess the rigor of our interventions.

The need for and value of standards for tests and measurements is clear—physical therapists must have a more scientific basis for practice. We test and measure in our daily practice, yet the validity and reliability of some of these tests may be questioned. With this document, we have guidelines to determine the quality of our tests and measurements.

Measurements are fundamental to the practice of physical therapy. We need to sit back and look at what we do each day, and strive for the quality we are capable of providing. Meaningful and useful measurements are important if we, as physical therapists, are to be recognized as credible health care providers.

Achieving a high quality of physical therapy practice requires us to evaluate the client, selecting and administering a variety of tests and measurements. We take our findings, interpret the data, and establish a baseline for the client's status. We then develop plans for therapeutic intervention that will achieve the goals we have set for the client. But how objective and accurate are these findings? How reliable? How valid? Can we select the appropriate interventions if our assessments are in question?
Standards provide the foundation for assessment of the quality of our practice. We use a variety of quality assurance methodologies to determine the degree to which the standards are met, and we take actions to improve the care when standards are not met. Quality assurance is the responsibility of every physical therapist, as well as the responsibility of the profession as a whole. Quality assurance continues to be an evolving process. The tools may change, but the objective remains the same: to improve patient care.

The physical therapy evaluation is the foundation for the measurement of the outcome of our therapeutic intervention. And we must measure these outcomes. In the past, quality assurance activities have focused more on the structure and process of our services. With the spiraling cost of health care in the United States, we must demonstrate the effectiveness and efficiency of our treatment. Quality assurance studies with an outcome focus can provide a measure of our progress toward achieving that goal.

We have our Standards of Practice adopted by the House of Delegates of the American Physical Therapy Association, and these standards assist us in our quality assurance mechanisms. Now, we have the Standards for Tests and Measurements in Physical Therapy Practice to assist us in ensuring the quality of our physical therapy evaluation. Clinicians must take these criteria and try to incorporate them into their daily practice.

The Task Force on Standards for Measurement in Physical Therapy has completed a complex task in advancing our knowledge and has provided a cornerstone of objective, reliable, and standardized tests and measurements. The Task Force members are to be commended for their hard work.

Elizabeth Gaynor, MS, PT
Chairperson, Committee on Physical Therapy Practice

Dedication

The members of the American Physical Therapy Association's Task Force on Measurement dedicate this document to Dr Steven J Rose, who died before the document was completed. Dr Rose was a visionary within physical therapy. He saw the need for standards for tests and measurements and welcomed the creation of a task force. He gladly accepted a position on the task force, despite the fact that he was ill. During the early phases of writing this document, he displayed remarkable courage, overcoming pain and disability to attend meetings. He was a vigorous participant in discussions. We note with pride his remarks that his excitement about this project led him to work longer and harder than he thought he could and that, in the midst of task force business, he even forgot about his pain and fatigue. Dr Rose's brilliant mind, his penchant for playing the devil's advocate, and his commitment to excellence were missed in the latter stages of this project. His spirit, however, remains in his many contributions to the Standards and accompanying documents and in the way these documents attempt to combine science and practice, Dr Rose's two great loves.
Introduction

Examination of physical therapy practice demonstrates the growing importance of measurement. Walking through a physical therapy clinic, you may observe a patient's range of motion being measured, or you may see a therapist testing the inspiratory capacity of a patient. Other therapists may be measuring the developmental status of a child or the accessory motion of the knee joint in a postsurgical patient. Still other therapists may be measuring the functional status of a patient with hemiplegia. Physical therapists need to obtain measurements because they make decisions, offer consultative opinions, and document changes in patient status.

This document, Standards for Tests and Measurements in Physical Therapy Practice, has been prepared because of the growing importance of measurement in physical therapy. Measurements are taken to provide information, but the result may be misinformation if the quality of measurements is not ensured. The purpose of this document is to provide standards that will help ensure the quality of measurements. These Standards are tools for practitioners. They are designed to provide guidelines that practitioners can use when they take measurements. The Standards are meant to represent the best in measurement and are not intended to hinder practice by establishing rigid rules that interfere with patient care. The Standards also demonstrate to society the commitment of physical therapists to practice in a credible and scientific manner. The Standards reflect our profession's humanistic commitment to provide the highest quality of care to our patients. The Standards include a section on research. The Standards, however, are primarily related to practice. They set how measurements should be used in clinical practice. Through the use of the Standards, therapists can, in their practice settings, deliver more effective care and document the results of treatment.

As clinicians, we cannot practice unless we take measurements. We need measurements in order to classify and describe patients, plan treatments, predict outcomes, document the results of treatments, determine the effectiveness of treatments, and determine when to refer patients to other practitioners. We may wonder how much more effectively we might practice if we knew more about our measurements, for example, if we knew when we should rely on our measurements and when we should seek confirming information. In addition to needing measurements for decision making, we need measurements in order to document what we are doing. In the face of shrinking resources for health care, society is no longer willing to accept on good faith alone the benefit of what we physical therapists do for our patients. Even widely accepted treatments may, in the future, become suspect if the measurements that justify these treatments are shown to be questionable.

This is an age when documentation, efficacy, and cost-effectiveness are increasingly important to those who control the reimbursement for all of health care, including physical therapy. Measurement will play an increasing role in determining who gets paid for doing what to whom, and for how long. Documentation with measurements of high quality may be the only way we physical therapists can ensure that our services will be available to persons who need these services. Physical therapy as a form of health care is at risk unless the results of physical therapy are judged to be worthwhile, not only by physical therapists and consumers of physical therapy, but also by third-party payers and corporate-world purchasers of health care. Some or much of what is being done in physical therapy could be denied reimbursement if we do not satisfactorily document the efficacy and cost-effectiveness of treatment. Without such reimbursement, physical therapy services could be denied to the very people who need our services the most. Proper attention to the quality of measurement in clinical practice will, therefore, not only ensure our profession's continued growth but also protect our patients.

Growth in the profession of physical therapy has taken place even though our profession has had no accepted standards for measurement and despite the fact that few education programs have prepared new therapists to understand what constitutes good measurement. Continued growth cannot be ensured unless the state of our measurements changes—and unless it changes soon. In March of 1986, the Board of Directors of the American Physical Therapy Association (APTA) recognized the need to improve the state of measurement in physical therapy. The Board made improved measurement a major goal of the Association.

Because resources in clinical measurement were limited, the Board called upon the Research Committee to develop a proposal for the development of Standards for Measurement in Physical Therapy. In August of 1987, the Research Committee, after consulting with experts on clinical measurement, developed a proposal for the development of Standards for Measurement in Physical Therapy. In November of 1987, the Board funded the proposal and made a commitment to a 2-year effort that would culminate in the publication of these Standards. The Task Force on Standards for Measurement in Physical Therapy was appointed to carry out this mission.

Physical therapy is not the first profession to recognize the need to improve the quality of its measurements. The American Psychological Association (APA) has been publishing monographs on standards in testing since 1954. The American Educational Research Association and the National Council on Measurement in Education joined forces with the APA in the mid-1960s and formed a joint committee that wrote Standards for Educational and Psychological Tests and Manuals. A 1966 version of the Stan-
One of the first acts of the Task Force members was to examine the APA manual. The APA Standards are a primary source of information on measurement. The members of the Task Force agreed, however, that while the APA document contained a great deal of useful information, it was not directly applicable to physical therapy. Many of the measurement problems in physical therapy are unique. Physical therapists use measurements that are based on the behavioral, biological, and physical sciences. The scope of measurements in physical therapy is extraordinary. Questionnaires are used by therapists, as are manual muscle testing, developmental testing, postural evaluations, instrumented muscle testing, movement analysis, and a whole variety of other tests. Instruments vary from paper and pencil, to the therapist’s hands, to complex computer-based machines with elaborate peripheral devices. We concluded that clinicians needed standards written by physical therapists for physical therapy practitioners. The Task Force, therefore, set out to develop Standards specifically for physical therapy. In developing these Standards, the Task Force was aware that most physical therapists receive little or no training in the science of measurement. The Task Force agreed that the final document must be sufficiently comprehensive to cover the vast expanse of physical therapy measurements and that it must also be practical.

The process of developing standards began with the entire Task Force considering philosophical and practical issues during 2 days of often heated, and always thorough, discussion. After the Task Force worked out basic concepts, the writing of the Standards was delegated to a three-member Working Group (Jules M Rothstein, Task Force coordinator; John L Echternach; and Harry G Knecht). The Working Group developed a draft document that was initially reviewed by the rest of the Task Force. The draft was revised. The present version has been revised on feedback from the physical therapy community, as well as on feedback from other interested parties and from experts on measurement.

In the training of physical therapists, measurement has all too often been equated with research. Concerns about the quality of measurements are, at times mistakenly, thought to relate to research and not to practice. Because therapists need help with the basic science of measurement, a primer on measurement has been prepared to complement the Standards. The purposes of the primer are to provide physical therapists with explanations of basic concepts and to explore issues related to measurement. Eugene Michels, who began work with the Task Force as an APTA staff member, wrote the initial draft of the primer. The primer is a tool that can be used to help understand not only the Standards, but also issues related to tests and measurements themselves. The primer is an independent, but complementary, document for the Standards.

The Standards include a glossary. The glossary defines terms as they are used in the Standards. The glossary allows readers to see how these terms were used by the authors of the Standards. The Task Force made every effort to avoid creating new terms and to avoid using jargon. The glossary is meant solely as a source for materials in the Standards; it is not a general measurement glossary. Wherever possible, the terms used and defined are those commonly found in the measurement literature. Clinicians may find many of the terms unfamiliar at first, but the Task Force believes that, through use of the glossary and the primer, the Standards can be understood and used by all therapists. The Task Force also recognizes that many physical therapists will have to make a commitment of time and effort to learn these new terms and to learn about measurement. In the future, these terms will be more commonplace in the clinical literature of physical therapy.

The Standards are meant to foster the continued growth of high-quality care in physical therapy. They are highly specific in describing what should be done to ensure meaningful and useful measurements. Part of the Standards provides long-overdue guidance to persons developing tests and to persons teaching about testing. No longer will clinicians independently have to ask purveyors of tests to supply vital information. The Standards specify what the providers should provide.

The Standards consist of five sections. The first three sections specify what is expected of test purveyors. Three categories of purveyors are described: primary purveyors, who originate tests; secondary purveyors, who conduct research and advocate the use of tests; and tertiary purveyors, who are teachers. The fourth section contains the Standards for Test Users—physical therapists. The fifth section describes standards for ensuring integrity in measurement studies. This last section is adapted from the APTA’s Standards for Integrity in Physical Therapy Research.

The Task Force originally had hoped to generate a series of guidelines that would be few in number and “user friendly.” Early versions of the Standards proved that this was impossible. Attempts to generate fewer sections and a more multipurpose document resulted in a cumbersome set of standards that was difficult to apply. Because this is the first document of its type in physical therapy and because of the nature of the subject, we found that the Standards needed to be comprehensive and to contain detailed specifications. We found that when we attempted to produce briefer versions of the Standards and when we attempted to use fewer measurement terms, our drafts were unclear and could not be used as references. The Standards, although they are advisory, may read like a rule book. Such books are not easily read, nor are they
commonly read from cover to cover. We chose to separate the Standards into sections, each for a specific audience. For example, physical therapists will usually be acting as test users and should read and consult the section designed for them. Therapists may, however, on occasion want to know what they should expect from purveyors of tests. When this need is recognized, they can consult the appropriate purveyor section.

In the Standards, the word "must" appears frequently. The Task Force consciously adopted the use of this word to provide a clear message about measurement. This message is that physical therapists who make decisions about measurements and their uses should understand that the use of the best measurements possible is obligatory. It is equally important to understand other ideas in this context. There is no intention to have anyone act as an enforcer of the Standards. The Standards represent an ideal; they represent a guide that therapists can use in their professional conduct. Measurements in physical therapy will improve when each therapist considers his or her own responsibility regarding the Standards.

The Standards provide a framework for professional decisions. They are optimal guidelines, not fixed, inviolate rules. Because standards are by their very nature statements of optimal characteristics, there is still considerable room for judgment.* Task Force members believe that measurements that fail to meet the Standards are less than ideal and that every effort should be made to avoid using such measurements. When that is not possible, clinical practice cannot wait and testing will usually have to proceed; however, physical therapists should be aware of and should acknowledge the limitations of the measurements they are using. The Standards should also heighten an awareness that the business of measurement should not be taken casually. The development of tests and measurements takes commitment and is often an arduous process that is marked by periods of testing and refinement. Therefore, although tests cannot necessarily meet all of the Standards, it is the responsibility of all persons promoting and using tests and measurements to make sure that reasonably acceptable adherence to the Standards occurs and that future efforts will be made at refinement.

This document is open to review. The published Standards for Tests and Measurements in Physical Therapy Practice represent only the beginning of an important effort. As the Standards continue to evolve, we hope that physical therapists will aid that endeavor by sharing with us their impressions and experiences with the Standards.

We believe that the Standards will become an essential part of physical therapy practice. Knowledge of measurement is no less important for clinical practice than is knowledge of anatomy, kinesiology, physiology, or psychology. All of these areas, including measurement, provide the scientific foundation for effective clinical practice. These are the profession's Standards, and, as such, they are a means of ensuring better care for our patients and of ensuring that physical therapists play an important role in the delivery of health care services.

*The Standards are not intended to codify, explain, modify, or replace any part of the ethical principles in the APTA's Code of Ethics or any of the interpretations in the Guide for Professional Conduct, which is issued by the Association's Judicial Committee.
Glossary of Terms Used in the Standards

The glossary describes terms as they are used in the Standards. The glossary is not meant to be all-inclusive, but rather to provide definitions for the terms as they are used in the Standards. For further information about the terms, related concepts, or other terms used in measurement, consult the Primer.

Alternate-forms (parallel-forms) reliability: see reliability

Assessment: measurement, quantification, or placing a value or label on something; assessment is often confused with evaluation; an assessment results from the act of assessing (see evaluation and examination)

Attribute: a variable; a characteristic or quality that is measured

Classification (categorization): assignment of an individual or an entity to a group; assignment is based on rules; groups are defined so that they allow all pertinent entities or individuals to belong to the defined groups (classes or categories are exhaustive) and so that they allow entities or individuals to belong to only one possible group (classes or categories are mutually exclusive)

Clinical decision: a determination that relates to direct patient care, indirect patient care, acceptance of patients for treatment, and whether patients should be referred to other practitioners (this definition is modified from that presented by Charles Magistro at a conference on Clinical Decision Making held under APTA auspices in October 1988 in Lake of the Ozarks, Missouri); a diagnosis that leads a therapist to take an action is a form of a clinical decision; clinical decisions result in actions; when direct supporting evidence for clinical decisions is lacking, such decisions are based on clinical opinions

Clinical opinion: a belief or idea that a physical therapist holds regarding a patient; this opinion may be based on the use of tests and measurements, but is not directly supported by evidence relating to those tests and measurements; clinical opinions are based on the therapist's evaluation of available information; clinical decisions (ie, determinations that cause the therapist to take an action) that are based on the therapist's synthesis of information are based on the clinical opinions of that therapist

Concurrent validity: see validity

Construct: a concept developed for the purpose of measurement; support for the construct is through logical argumentation based on the theoretical and research evidence (see construct validity listed under validity)

Construct validity: see validity

Content validity: see validity

Criterion-based (criterion-related) validity: see validity

Data: synonymous with measurements (see measurement)

Derived measurement: a measurement of an attribute that is obtained as the result of a mathematical operation applied to an existing measurement of some other attribute; an example is the measurement of leg-length difference, which is derived by subtracting one leg-length measurement from another

Evaluation: a judgment based on a measurement; often confused with assessment and examination (see assessment and examination); evaluations are judgments of the value or worth of something

Examination: a test or a group of tests used for the purpose of obtaining measurements or data (see assessment and evaluation)

False negatives: persons who test negatively for some attribute but who, in fact, have that attribute (see true negatives)

False positives: persons who test positively for some attribute but who, in fact, do not have that attribute (see true positives)

Instrument: a machine, a questionnaire, or any device that is used as part of, or as a test to obtain, measurements

Internal consistency: see reliability

Intertester reliability: see reliability

Intratester reliability: see reliability

Measure: the act of obtaining a measurement (datum)

Measurement: the numeral assigned to an object, event, or person or the class (category) to which an object, event, or person is assigned according to rules

Normalization: a process that yields a new or transformed measurement that is mathematically derived to change the distribution of measurements; normalization procedures are often used to change the distribution of data to make the distribution more congruent with a bell-shaped (or normal) curve

Objective measurement: a measurement that is not affected by some aspect of the person obtaining the measurement; the opposite of a subjective measurement (see subjective measurement); measurements cannot be
totally objective, because the term "objective" relates to the reliability of measurements, especially the intertester reliability; objectivity and reliability are measured along a continuum

**Operational definition:** a set of procedures that guides the process of obtaining a measurement; includes descriptions of the attribute that is to be measured, the conditions under which the measurement is to be taken, and the actions that are to be taken in order to obtain the measurement

**Parallel-forms (alternate-forms) reliability:** see reliability

**Practicality of a test:** the usefulness of a test based on issues relating to personnel, time, equipment, cost of administration, and impact on the person taking a test

**Predictive validity:** see validity

**Predictive value of a measurement:** the degree of certainty that can be associated with a positive or negative finding (measurement) obtained on a diagnostic test; the predictive value of a positive measurement is the ratio formed by dividing the number of true positives by the number of all positive findings; the predictive value of a negative measurement is the ratio formed by dividing the number of true negatives by the number of all negative findings

**Prescriptive validity:** see validity

**Primary purveyor:** see purveyor

**Purveyor:** any person (or organization) who develops a test or any person (or organization) who offers, promotes, or requires the use of a test; a purveyor is also a person who advocates use of specific tests through the publication of research or scholarly articles or through teaching

**Primary purveyor:** a person who develops, promotes, or requires the use of tests; this definition includes persons within clinical institutions who require the use of specific tests; persons who conduct continuing education courses in which a major component involves the advocacy of the use of specific testing procedures are primary purveyors; any person (or organization) who promotes (advocates) the use of tests by selling testing equipment, manuals, books, or similar materials is a primary purveyor; in the case of books or articles that serve as test manuals, the primary purveyor is the author; persons who sell instruments that may be used for testing, but who do not describe or advocate specific testing procedures, are not purveyors (see purveyor, secondary purveyor, and tertiary purveyor)

**Secondary purveyor:** any researcher or other person who publishes a scholarly work that examines aspects of tests and who, in that scholarly work, suggests (advocates) that a test be used; a secondary purveyor is not the initial source of information on a test (ie, did not supply the manual or the original information on the test) (see purveyor, primary purveyor, and tertiary purveyor)

**Tertiary purveyor:** any person who teaches or prepares instructional material that describes specific tests or specific uses of measurements; this definition includes, but is not limited to, persons teaching in academic institutions, clinical educators, and continuing educators who are not acting in the role of primary or secondary purveyors (see purveyor, primary purveyor, and secondary purveyor)

**Reactivity:** the degree to which the process of taking a test affects a measurement or other measurements taken on the same person in the future; examples are learning and physiological effects of taking tests

**Reliability:** the consistency or repeatability of measurements; the degree to which measurements are error-free and the degree to which repeated measurements will agree

**Internal consistency:** the extent to which items or elements that contribute to a measurement reflect one basic phenomenon or dimension

**Intertester reliability:** the consistency or equivalence of measurements when more than one person takes the measurements; indicates agreement of measurements taken by different examiners

**Intratester reliability:** the consistency or equivalence of measurements when one person takes repeated measurements separated in time; indicates agreement in measurements over time

**Parallel-forms (alternate-forms) reliability:** the consistency or agreement of measurements obtained with different (alternative) forms of a test; indicates whether measurements obtained with different forms of a test can be used interchangeably

**Test-retest reliability:** the consistency of repeated measurements separated in time; indicates stability (reliability) over time

**Score (grade):** the numeric (quantitative) or verbal (qualitative) descriptor used to characterize the result of a test; a score is a measurement (see measurement)

**Secondary purveyor:** see purveyor
Sensitivity of a test: an indication of how well a diagnostic test identifies people who should have a positive finding; the numerical representation of sensitivity is a ratio formed by dividing the number of persons with a true-positive response on a test by the number of persons who should have had a positive response (ie, the number of persons who are known to have properties that would indicate that they should test positive).

Specificity of a test: an indication of how well a diagnostic test identifies people who should have a negative finding; the numerical representation of specificity is a ratio formed by dividing the number of persons with a true-negative response on a test by the number of persons who should have had a negative response (ie, the number of persons who are known to have properties that would indicate that they should test negative).

Standardization: a process by which a score is converted (transformed) into a relative score by using indices of central tendency and variability; a commonly used standardized score is the z score; the term "standardization" is also used to describe the process of systematization of the methods used to obtain a measurement; the process of standardization, however, does not ensure reliability, because reliability can only be determined through the collection of data (see reliability).

Subjective measurement: a measurement that is affected by some aspect of the person obtaining the measurement (contrasts with objective measurement); subjectivity relates to the reliability of measurements, especially the intertester reliability; the more subjective the measurement, the less reliable the measurement; subjectivity, like reliability, is measured along a continuum.

Tertiary purveyor: see purveyor.

Test: a procedure or set of procedures that is used to obtain measurements (data); the procedures may require the use of instruments.

Test manual: a booklet or book prepared by a primary test purveyor to guide the process of obtaining a measurement and to provide documentation and justification for the test.

Test setting: the environment in which a test is given, including the physical setting and the characteristics of that setting.

Test user: one who chooses tests, interprets test scores, or makes decisions based on test scores (this definition is from Standards for Educational and Psychological Tests; American Psychological Association, Washington, DC, 1974, page 1).

Test-retest reliability: see reliability.

Transformation of measurements: the application of a mathematical operation for the purpose of changing the value or distribution of measurements, such as is done in the process of standardization or normalization.

True negatives: persons who test negatively for some attribute and who, in fact, do not have that attribute (see false negatives).

True positives: persons who test positively for some attribute and who, in fact, have that attribute (see false positives).

Validity: the degree to which a useful (meaningful) interpretation can be inferred from a measurement.

Concurrent validity: a form of criterion-based validity in which an inferred interpretation is justified by comparing a measurement with supporting evidence that was obtained at approximately the same time as the measurement being validated.

Construct validity: the conceptual (theoretical) basis for using a measurement to make an inferred interpretation; evidence for construct validity is through logical argumentation based on theoretical and research evidence (see construct).

Content validity: a form of validity that deals with the extent to which a measurement is judged to reflect the meaningful elements of a construct and not any extraneous elements.

Criterion-based (criterion-related) validity: three forms of criterion-based validity exist: concurrent validity, predictive validity, and prescriptive validity; the common element is that, with each of these forms of validity, the correctness of an inferred interpretation can be tested by comparing a measurement with either a different measurement or data obtained by other forms of testing.

Predictive validity: a form of criterion-based validity in which an inferred interpretation is justified by comparing a measurement with supporting evidence that is obtained at a later point in time; examines the justification of using a measurement to say something about future events or conditions.

Prescriptive validity: a form of criterion-based validity in which the inferred interpretation of a measurement is the determination of the form of treatment a person is to receive; prescriptive validity is justified based on the successful outcome of the chosen treatment.
Organization of the Standards for Primary Purveyors:
Primary purveyors are obliged to provide documentation of essential elements for the tests and measurements they are promoting. Documentation should be in the form of a test manual. Most of the Standards for primary purveyors describe the elements that should be included in test manuals. Qualitative requirements for the information to be included in the manuals are presented within sections that describe what should be included in the test manuals.

P1. Persons or organizations should not become primary test purveyors unless they are prepared to adhere to the Standards.

P2. Primary purveyors of tests must provide test manuals. Books that contain major sections dealing with tests and include materials that promote and advocate the use of tests are considered test manuals, and all standards for test manuals apply to these books. Primary purveyors are responsible for the quality (accuracy) of all information in their manuals and must make every effort to ensure that information in the manuals is in compliance with the Standards (eg, research studies cited are in accordance with the Standards).

P3. Test manuals provided by primary purveyors must include descriptions of the theoretical bases of the tests and measurements, including discussions of the evidence supporting the construct validity and the content validity of the measurements. The purpose of the test must be clearly described.

P4. Test manuals provided by primary purveyors must include operational definitions.

P4.1. Operational definitions of attributes that the test measures must be provided in the test manual.

P4.2. Operational definitions of terms used to describe the population for whom the test is intended must be provided in the test manual.

P4.3. Operational definitions of terms used to describe potential test users must be provided in the test manual.

P4.4. Operational definitions of terms used to describe components of the test or test instruments must be provided in the test manual.

P4.5. Operational definitions of any unique terms created by the primary purveyor must be provided in the test manual.

P4.6. Operational definitions of any terms used in a noncustomary (unusual) manner by the primary purveyor must be provided in the test manual.

P5. Test manuals provided by primary purveyors must include descriptions of the populations for whom the tests are designed. Descriptions of subjects for whom the tests should not be used and descriptions of subjects for whom the tests should be used with caution must be included.

P6. Test manuals provided by primary purveyors must include descriptions of procedures that will ensure safe test administration. Safety procedures must be enumerated and should include specific instructions as to when the test should be terminated if a subject has an adverse response.

P7. Test manuals provided by primary purveyors must include descriptions of the qualifications and competencies needed by test users. These descriptions should include statements regarding potential consequences of unqualified users administering the test.

P8. Test manuals provided by primary purveyors should describe how potential test users can obtain the competencies necessary to administer the tests.

P9. Test manuals provided by primary purveyors should include narrative chronological accounts of the development of the tests, including descriptions of the development of any instruments associated with the tests.

P9.1. A description of the test developer(s) must be provided in the narrative account in the test manual.
P9.2. A description of the setting(s) in which the test was developed must be provided in the test manual.

P9.3. Documentation of the sources for any items, components, or elements used in the test must be provided in the narrative account in the test manual.

P9.4. A summary description of the history of the test, including where and how the test has been used, must be provided in the narrative account in the test manual.

P9.5. Descriptions of any revisions of the test and explanations of why revisions were made in the test must be provided in the narrative account in the test manual.

P10. Test manuals provided by primary purveyors must include descriptions of the test and associated instruments.

P10.1. Documentation of relevant technical information regarding performance characteristics of any machines, recording devices, transducers, computer interfaces, and similar instruments must be provided in the test manual.

P10.2. Descriptions of how instruments used in the test manipulate or process information in order to obtain the desired measurements must be provided in the test manual.

P11. Test manuals provided by primary purveyors must include instructions for administering the tests described in the manual. These instructions must include descriptions of all equipment and activities needed for obtaining, recording, interpreting, and reporting the measurements.

P11.1. Guidelines must be provided in the test manual as to what information and instructions should be given to the person being tested. In order to allow test users to answer questions about the test and related topics, adequate information about the test should be provided in the test manual.

P11.2. Guidelines should be provided in the test manual as to what actions persons administering the test can take to minimize the effects of extraneous factors on test performance.

P11.3. Descriptions must be provided in the test manual of the physical settings in which tests should be given and the possible effects of conducting the test in other settings.

P11.4. Descriptions must be provided in the test manual of test conditions, behaviors of persons taking the test, and other factors that could make the validity of the measurements questionable.

P11.5. Descriptions must be provided in the test manual of how the test user must manipulate or process information in order to obtain the desired measurements.

P11.6. Descriptions and instructions must be provided in the test manual for the use of any instruments required to obtain the desired measurements. This information must include, where appropriate, machine settings and any other user-selected options. The test manual must include descriptions of the effects of all options on the measurements and the consequences of selecting the incorrect options.

P11.7. If instruments are used as part of the test, the test manual must include descriptions of how the devices are calibrated. A means of testing calibration must be described in the test manual. If calibration is needed, instructions must be provided regarding a course of action to be taken.

P11.8. Descriptions must be provided in the test manual of variations in the test procedures that are available to the test user. Descriptions of variations that are known not to impair the quality of the measurements and descriptions of variations that are known to lead to measurements of questionable validity must be included.

P11.9. Background information must be provided in the test manual so that test users have the knowledge to obtain any derived measurements or categorizations necessary for interpretation of the measurements.

P11.10. Warnings must be provided in the test manual regarding the misuse of the measurements. Common errors in interpretation of the obtained measurements must be described.

P12. Test manuals provided by primary purveyors must include discussions of reactivity.

P12.1. Discussion of the degree to which administration of the test affects the measurements obtained from that test or any subsequent tests must be provided in the test manual.

P12.2. Discussion of the degree to which administration of the test may cause a change in the person taking the test must be provided in the test manual. Discussions of side effects, aftereffects,
and the effects of fatigue, learning, pain, and so forth may be included.

P13. Test manuals provided by primary purveyors must include evidence for all relevant forms of reliability and related information for the measurements described in the test manual.

P13.1. Descriptions of how information related to reliability was collected must be provided in the test manual, and all relevant references to peer-reviewed publications must be supplied.

P13.2. Evidence relating to reliability must be reported in the test manual in a way that describes the errors associated with common uses of the measurements.

P13.2.1. Intratester reliability estimates (indices) must be reported in the test manual. Within-day and between-day studies should have been conducted in a clinical context consistent with the intended use of the measurements. Intratester reliability should be reported in the test manual for all forms of measurements, including self-administered tests.

P13.2.2. Intertester reliability estimates (indices) must be reported in the test manual. Intertester reliability studies should have been conducted in a clinical context consistent with the intended use of the measurements.

P13.2.3. Internal consistency coefficients (or factor structures) must be reported in the test manual when there is a need to demonstrate that items or elements contributing to a measurement reflect one basic phenomenon or dimension. Studies of internal consistency should have been conducted in a clinical context consistent with the intended use of the measurements.

P13.2.4. Parallel-forms (alternative-forms) reliability must be reported in the test manual if more than one version of the test is being described. Studies of parallel-forms reliability should have been conducted in a clinical context consistent with the intended use of the measurements.

P14. Test manuals provided by primary purveyors should include descriptions of all research studies into the reliability of the measurements described in the manual, and all relevant references to peer-reviewed publications must be supplied.

P14.1. Descriptions of who conducted the reliability research must be provided in the test manual.

P14.2. Descriptions of where the reliability research was conducted must be provided in the test manual.

P14.3. Descriptions of the sample(s) studied in the reliability research must be provided in the test manual.

P14.3.1. Descriptions must be provided in the test manual of how the sample studied in the reliability research was selected.

P14.3.2. The number of subjects studied in the reliability research must be specified in the test manual.

P14.3.3. Descriptions of relevant clinical characteristics of the sample studied in the reliability research must be provided in the test manual. A discussion of how the sample is representative of the population for whom the test is intended should be included in the test manual.

P14.4. Descriptions of persons who obtained the measurements in the reliability research (ie, those who were in the role of test users) must be provided in the test manual. Descriptions of their qualifications, competencies, and experiences with the test should be included. Any special information or training given to test users prior to their obtaining the measurements in the study should be described in the test manual.

P14.5. Descriptions of the methods and research design used in the reliability studies must be provided in the test manual. The specific types of reliability that were investigated must be specified.

P14.6. Descriptions of the statistics used to derive reliability estimates and the rationale for their use must be provided in the test manual. When methodologically appropriate, reports of confidence intervals and standard errors of measurements should be included in the test manual. Examples of how the reliability estimates are to be used as part of data interpretation should be included. Reliability estimates should be accompanied by reports of regression data (ie, slopes and intercepts) when appropriate for the statistical analysis.

P15. Test manuals provided by primary purveyors must include evidence for all relevant forms of validity
and related information for the measurements described in the manuals. All relevant references to peer-reviewed publications must be supplied.

P15.1. Descriptions of how information related to validity was collected must be provided in the test manual, and references to all relevant peer-reviewed publications must be supplied in the test manual.

P15.2. Evidence relating to validity must be reported in the test manual in a way that describes the errors associated with common uses of the measurements.

P15.2.1. The construct validity (theoretical basis) for the use of the measurement must be discussed in the test manual. Experimental evidence as well as logical arguments for the intended use of the measurements should be provided in the test manual.

P15.2.2. The content validity of the measurements must be discussed in the test manual. Experimental evidence as well as logical arguments for the content validity of the measurements should be provided in the test manual.

P15.2.3. Evidence for concurrent validity must be provided in the test manual when the primary purveyor contends that the measurements can be used to make inferences about the current status of an attribute at the time the measurements are obtained or shortly thereafter. This evidence must include logical and experimental data to support the use of other measurements as criteria to justify a concurrent inference. The primary purveyor should not make claims in the test manual for concurrent validity by comparing the measurement of interest with another measurement (the criterion) unless the criterion measurement has been shown to be valid (ie, it has been justified for use as a criterion).

P15.2.4. Evidence for predictive validity must be provided in the test manual when the primary purveyor contends that the measurements can be used at the time they are obtained to make inferences about the future status of an attribute. This evidence must include logical and experimental data to support the use of other measurements as criteria to justify a predictive inference. The primary purveyor should not make claims in the test manual for predictive validity by comparing the measurement of interest with another measurement (the criterion) unless the criterion measurement has been shown to be valid (ie, it has been justified for use as a criterion).

P15.2.5. Evidence for prescriptive validity must be provided in the test manual when the primary purveyor contends that the measurements can be used to determine the choice of treatment. This evidence must be based on research indicating that treatment chosen on the basis of the measurement is effective. Documentation of the effectiveness of treatment in the test manual must be based on the use of valid measurements.

P16. Test manuals provided by primary purveyors should include descriptions of all research studies into the validity of the measurements (see standard P15 for details on requirements for validity studies).

P16.1. Descriptions of who conducted the validity research must be provided in the test manual.

P16.2. Descriptions of where the validity research was conducted must be provided in the test manual.

P16.3. Descriptions of the sample(s) studied in the validity research must be provided in the test manual.

P16.3.1. Descriptions of how the sample in the validity research was selected must be provided in the test manual.

P16.3.2. The number of subjects studied in the validity research must be specified in the test manual.

P16.3.3. Descriptions of relevant clinical characteristics of the sample studied in the validity research must be specified in the test manual.

P16.4. Descriptions of persons who obtained the measurements in the validity research (ie, those who were in the role of test users) must be provided in the test manual. Descriptions of their qualifications, competencies, and experiences with the test should be included. Any special information or training given to test users prior to their obtaining the measurements in the study should be described in the test manual.
P16.5. Descriptions of the methods and research design used in the validity studies must be provided in the test manual. The specific types of validity that were investigated must be specified in the test manual.

P16.6. Descriptions of the statistics used to derive validity estimates and the rationale for their use must be provided in the test manual. Examples of how the validity estimates are to be used as part of data interpretation should be included in the test manual. Reports of estimates of validity in the test manual should be accompanied by reports of regression data (ie, slopes and intercepts) and the standard error of the estimate when methodologically appropriate.

P17. Primary purveyors who claim that measurements can be used to classify persons into diagnostic groups based on the presence or absence of a finding (eg, cut scores or tests that result in determinations of negative or positive findings) must include in their test manuals the essential elements that allow for interpretation of these findings. In reporting these elements, the same standards as described for reports of validity must be followed.

P17.1. Percentages of false positives and false negatives must be reported in the test manual.

P17.2. Sensitivity of the test must be reported in the test manual.

P17.3. Specificity of the test must be reported in the test manual.

P17.4. Predictive values of positive and negative findings (measurements) obtained with the test must be reported in the test manual.

P18. Test manuals provided by primary purveyors must include normative data when measurements are to be interpreted in terms of how an individual measurement compares with measurements obtained on other persons (ie, when the data are norm-referenced).

P18.1. Descriptions of who obtained the normative data must be provided in the test manual.

P18.2. Descriptions of where the normative data were obtained must be provided in the test manual.

P18.3. Descriptions of the sample studied to obtain the normative data must be provided in the test manual.

P18.4. Descriptions of persons who took the measurements used to obtain the normative data (ie, those who were in the role of test users) must be provided in the test manual. The test manual should include descriptions of test users’ qualifications, competencies, and experiences with the test. Any special information or training given to test users prior to their taking the measurements in the study should be described in the test manual.

P18.5. Descriptions of the methods and research design used to obtain the normative data must be provided in the test manual. Normative data should be obtained using the same measurement procedures that are described in the manual. If other versions of the test were used to obtain the normative data, or if other scales were used, there must be a discussion of how the normative data relate to the data that can be obtained using the test described in the manual.

P18.6. A complete discussion of limitations in the use of the supplied normative data must be provided in the test manual. The discussion may include, but should not be limited to, considerations of whether the normative data relate to a particular local area, facility, ethnic group, age group, or gender.

P18.7. Details on any data transformations (eg, any standardization or normalization procedures) used in obtaining or preparing the normative data must be provided in the test manual.
P18.8. Primary purveyors who describe measurements that are based on interval or ratio scales should present in the test manual as part of the normative data standard scores or percentiles with accompanying measures of central tendency and variability. Data for clinically meaningful subgroups should be similarly reported in the test manual.

P18.9. Primary purveyors who describe measurements that are based on ordinal or nominal scales should present in the test manual normative data in the form of the proportion of persons in the population who can be expected to belong to each group and subgroup. Data for clinically meaningful subgroups should be similarly reported in the test manual.

P19. Test manuals provided by primary purveyors must include information that will enable a user to judge the practicality of obtaining the measurements.

P19.1. Descriptions of the number and types of personnel needed to administer the test must be provided in the test manual.

P19.2. Estimates of the time required to administer the test should be provided in the test manual.

P19.3. Descriptions of any additional equipment or supplies needed to obtain the measurements should be provided in the test manual.

P19.4. Descriptions of any potential impact on the person taking the test, in terms of the person’s time and effort required and any other special requirements, should be provided in the test manual.

P19.5. Descriptions of any potential risks or hazards, and means for reducing the risks and hazards, to persons taking the test or administering the test must be provided in the test manual.

P20. Test manuals provided by primary purveyors must include discussions of any special considerations concerning the test and resulting measurements. For example, subgroups for whom measurements may be invalid should be identified, as should persons for whom a test may represent some psychological or physical risk.

P21. Test manuals provided by primary purveyors must include descriptions of all special groups for whom the test is contraindicated or known to lead to measurements of questionable validity.

P22. Test manuals provided by primary purveyors must include descriptions of a mechanism by which test users can communicate with the primary purveyor regarding the test. The mechanism should allow the user to seek further information, share observations and results, or report problems.

P23. Test manuals provided by primary purveyors must include a bibliography that provides references specific to the test and pertinent to the content of the test. The bibliography must be organized in such a manner that references supporting the scientific basis for the test are differentiated from references dealing tangentially with the test or simply reporting that the test has been used.

P24. A primary purveyor’s promotional material for a test or measurement must not make claims that exceed what can be justified by existing research. When information about a test or measurement is provided in promotional material, that material should meet the same standards of accuracy and freedom from misleading impressions that apply to the test manual (this standard is based on that found in the Standards for Educational and Psychological Tests: American Psychological Association, Washington, DC, 1974, page 10). Any primary purveyor who chooses to discuss the tests or measurements of another purveyor in promotional materials should do so only while maintaining the same standards of accuracy and freedom from misleading impressions that apply to the tests manual.

P25. Primary purveyors should make every reasonable effort to notify test users and potential users of any modifications or revisions in tests or test manuals.

P26. Primary purveyors who administer tests as part of test development must make every reasonable effort to observe the Standards for Test Users.
Standards for Secondary Test Purveyors (indicated with an S)

The Standards in this section describe requirements for secondary purveyors of tests. The following is the definition of a secondary purveyor.

Secondary purveyor: any researcher or other person who publishes a scholarly work that examines aspects of tests and who, in that scholarly work, suggests (advocates) that a test be used; a secondary purveyor is not the initial source of information on a test (i.e., did not supply the manual or the original information on the test) (see purveyor, primary purveyor, and tertiary purveyor)

Organization of the Standards for Secondary Purveyors: Secondary purveyors may have limitations imposed upon them regarding the information they can supply in published reports. The Standards have been written in a way that takes into account these limitations. However, the Standards do list the elements that should be included in written materials prepared by secondary purveyors. Using the Standards as a guide, secondary purveyors may, when publication limitations are too stringent, have to decide whether the integrity of their reports may have to be excessively compromised by the requirements for publication. Secondary purveyors are obligated to reconsider whether, in the face of such limitations, they choose to remain secondary purveyors.

S1. Persons or organizations should not become secondary test purveyors (i.e., advocates of using tests) unless they are prepared to adhere to the Standards. A scholarly publication that describes tests or uses of tests does not make an author a test purveyor unless advocacy of specific test use is part of that publication. Care should be taken in such publications to differentiate analysis, research, and discussion from advocacy.

S2. Secondary purveyors who advocate the use of tests or measurements must be prepared to support that advocacy. In journal articles, the advocacy should not exceed what can be supported through documentation. Secondary purveyors, therefore, should be aware of the limitations imposed by journals publishing their reports. Secondary purveyors who publish in other forums or who are not presenting a research report should attempt to supply more information than can be expected in a typical published research report.

S3. Secondary purveyors must include in all their research reports or scholarly papers the basic elements that will ensure credibility. Secondary purveyors should make every reasonable effort to publish their reports in peer-reviewed journals, which ordinarily require the basic elements for credibility.

S4. Secondary purveyors must include in all of their research reports or scholarly papers sufficient detail to allow for replication of their research.

S5. Secondary purveyors must include descriptions of the theoretical bases for the test and measurements they discuss in research reports or scholarly papers. A discussion of the evidence relating to the construct validity and the content validity of the measurements should be included. The purpose of the test must be clearly described. The length of these discussions should be to the extent allowed in the publication in which the report will or may appear.

S6. Secondary purveyors should include, to the extent allowed in the publication in which their report will or may appear, operational definitions related to all aspects of the tests and measurements they discuss.

S6.1. Operational definitions of attributes that the test measures must be provided in reports by secondary purveyors.

S6.2. Operational definitions must be provided for terms used to describe the population for whom the test is intended in reports by secondary purveyors.

S6.3. Operational definitions of terms used to describe potential test users must be provided in reports by secondary purveyors.

S6.4. Operational definitions of terms used to describe components of the test or test instruments must be provided in reports by secondary purveyors.

S6.5. Operational definitions of any unique terms created by the secondary purveyor must be provided in reports by secondary purveyors.

S6.6. Operational definitions of any terms used in a noncustomary (unusual) manner by the secondary purveyor must be provided in reports by secondary purveyors.

S7. Research reports or scholarly papers written by secondary purveyors must include, to the extent allowed in the publication in which the report will or may appear, a description of the population for whom the test is designed. Descriptions, based on research of the secondary purveyor, of subjects for whom the test should not be used and descriptions of subjects for whom the test should be used with caution should be included.

S8. Research reports or scholarly papers written by secondary purveyors should include, to the extent allowed in the publication in which the report will or may appear, descriptions of the qualifications and
competencies of persons who use (administer) the tests being discussed.

S9. Research reports or scholarly papers written by secondary purveyors should include, to the extent allowed in the publication in which the report will or may appear, a brief account of the development of the test being discussed.

S10. Research reports or scholarly papers written by secondary purveyors must include, to the extent allowed in the publication in which the report will or may appear, a description of the test being discussed and associated instruments.

S10.1. Documentation of relevant technical information regarding performance characteristics of any machines, recording devices, transducers, computer interfaces, and similar instruments must be provided in reports by secondary purveyors. Regardless of whether this information is published, a secondary purveyor must, upon request, be prepared to provide this information by personal communication.

S10.2. Research reports or scholarly papers written by secondary purveyors must include descriptions of how instruments manipulate or process information in order to obtain the measurements being discussed.

S11. Research reports or scholarly papers written by secondary purveyors must include instructions for conducting the test being discussed. These instructions must include, to the extent allowed in the publication in which the report will or may appear, descriptions of all activities needed for obtaining measurements, for recording measurements, and for interpreting measurements. Regardless of whether this information is published, a secondary purveyor must, upon request, be prepared to provide this information by personal communication.

S11.1. Research reports or scholarly papers written by secondary purveyors should include guidelines on the information that is to be given to the persons being tested. These guidelines should include the instructions given to the persons being tested.

S11.2. Research reports or scholarly papers written by secondary purveyors should include descriptions of the physical settings in which the test should be given and the possible effects of conducting the test in other settings.

S11.3. Research reports or scholarly papers written by secondary purveyors should include descriptions of test conditions, behaviors of persons taking the test, and other factors that could make the validity of the measurements questionable.

S11.4. Research reports or scholarly papers written by secondary purveyors should include descriptions of how the data were manipulated or processed in order to obtain the measurement being discussed.

S11.5. Research reports or scholarly papers written by secondary purveyors should include descriptions and instructions of how instruments were used to obtain the measurements being discussed. This information must include, where appropriate, machine settings and any other user-selected options. A discussion of the possible effects of any option on the measurements and the consequences of selecting the incorrect options should be included.

S11.6. Secondary purveyors who describe the use of instruments in their reports should include descriptions of how the instruments used to obtain the measurements are calibrated.

S11.7. Research reports or scholarly papers written by secondary purveyors should include sufficient background information so that readers can understand how any derived measurements or categorizations were made, especially if this information is necessary for interpretation of the measurements.

S11.8. Research reports or scholarly papers written by secondary purveyors must include warnings regarding misuse of the measurements being discussed. If research indicates that common errors in interpretation of test data can occur, secondary purveyors must describe these errors in their reports.

S12. Research reports or scholarly papers written by secondary purveyors should include, to the extent allowed in the publication in which the report will or may appear, discussions of special considerations concerning the test and measurements being discussed.

S12.1. Secondary purveyors should include discussions of reactivity in their reports.

S12.1.1. Research reports or scholarly papers written by secondary purveyors should include discussions of the degree to which administration of the test being discussed affects the measurement obtained from that test or any subsequent tests.
S12.1.2. Research reports or scholarly papers written by secondary purveyors should include discussions of the degree to which administration of the test being discussed may cause a change in the person taking the test. Discussions of side effects, after effects, and the effects of fatigue, learning, pain, and so forth may be included.

S13. Secondary purveyors who author research reports or scholarly papers that examine reliability must include in those reports, to the extent allowed in the publication in which the report will or may appear, essential elements that would allow for interpretation of the report.

S13.1. Research reports or scholarly papers on reliability written by secondary purveyors should include a thorough and critical review of what is known about the reliability and the validity of the measurements being discussed.

S13.2. Research reports or scholarly papers on reliability written by secondary purveyors must include a detailed description of the sample studied.

S13.2.1. Research reports or scholarly papers on reliability written by secondary purveyors must include descriptions of how the sample studied in the reliability research was selected.

S13.2.2. Research reports or scholarly papers on reliability written by secondary purveyors must specify the number of subjects studied.

S13.2.3. Research reports or scholarly papers on reliability written by secondary purveyors must include descriptions of relevant clinical characteristics of the sample studied. A discussion of how the sample is representative of the population for whom the test is intended should be provided by the secondary purveyor.

S13.3. Research reports or scholarly papers written by secondary purveyors must include descriptions of persons who obtained the measurements in reliability studies (ie, those who were in the role of test users). Descriptions of the test users’ qualifications, competencies, and experiences with the test should be included. Information or special training given to test users prior to their obtaining the measurements in the study should be described by the secondary purveyor.

S13.4. Research reports on reliability written by secondary purveyors must include descriptions of the methods and research designs used in their studies. The specific types of reliability being investigated must be specified.

S13.5. Research reports on reliability written by secondary purveyors must include a description of the statistics used to derive reliability estimates. The rationale for the use of these statistics must be provided. When methodologically appropriate, reports of confidence intervals and standard errors of measurement should be included. Examples of how the reliability estimates are to be used as part of data interpretation should be included. A reliability estimate should be accompanied by a report of regression data (ie, slopes and intercepts) when appropriate for the statistical analysis.

S14. Secondary purveyors who author research reports or scholarly papers that examine validity must include in those reports, to the extent allowed in the publication in which the report will or may appear, elements that allow for interpretation of the report.

S14.1. Research reports or scholarly papers on validity written by secondary purveyors should include a thorough and critical review of what is known about the reliability and the validity of the measurements being discussed.

S14.2. Research reports on validity written by secondary purveyors must include descriptions of the methods and research designs used in their studies. The specific types of validity investigated must be specified by the secondary purveyor. Descriptions of the sample(s) studied in the validity research must be provided. These descriptions should include the number of subjects studied and how these subjects were selected.

S14.3. Research reports or scholarly papers written by secondary purveyors must include evidence of validity to support each inferential use of the measurement suggested by the secondary purveyor. The design must be appropriate to support arguments for the presence of each relevant type of validity.

S14.4. Research reports on validity written by secondary purveyors must include descriptions of the statistics used to derive validity estimates. The rationale for the use of these statistics must be provided. When methodologically appropriate, reports of confidence intervals and standard errors of the estimate should be included. Examples of how the validity estimates are to be used as part of data interpretation should be included. Validity estimates should be accompanied by reports of regression data (ie, slopes
and intercepts) when appropriate for the statistical analysis.

S14.5. Secondary purveyors who state in research reports or scholarly papers that measurements can be used to make inferences about the current status of an attribute at the time the measurements are obtained or shortly thereafter must include logical and experimental data to support the use of other measurements as criteria to justify these concurrent inferences. Secondary purveyors should not make claims for concurrent validity by comparing the measurement of interest with another measurement (the criterion) unless the criterion measurement has been shown to be valid (ie, it has been justified for use as a criterion).

S14.6. Secondary purveyors who state in research reports or scholarly papers that measurements can be used at the time they are obtained to make inferences about the future status of an attribute must include logical and experimental data to support the use of other measurements as criteria to justify these predictive inferences. Secondary purveyors should not make claims for predictive validity by comparing the measurement of interest with another measurement (the criterion) unless the criterion measurement has been shown to be valid (ie, it has been justified for use as a criterion).

S14.7. Secondary purveyors who state in research reports or scholarly papers that measurements can be used to determine the choice of treatment (ie, prescriptive validity) must base these statements on research indicating that treatment chosen on the basis of the measurement is effective. Documentation of the effectiveness of treatment must be based on the use of valid measurements and should be included in reports by the secondary purveyors.

S15. Secondary purveyors who claim in research reports or scholarly papers that measurements can be used to classify persons into diagnostic groups based on the presence or absence of a finding (eg, cut scores or tests that result in determinations of negative or positive findings) must report the essential elements that allow for interpretation of these findings. In reporting these elements, the same standards as described for reports of validity must be followed. This information should be supplied to the extent allowed in the publication in which the report will or may appear.

S15.1. Percentages of false positives and false negatives must be described in reports by secondary purveyors.

S15.2. Sensitivity of the test must be described in reports by secondary purveyors.

S15.3. Specificity of the test must be described in reports by secondary purveyors.

S15.4. Predictive values of positive and negative findings (measurements) obtained with the test must be described in reports by secondary purveyors.

S16. Secondary purveyors who include normative data in their reports must include, to the extent allowed in the publication in which the report will or may appear, essential elements required for the interpretation of these normative data.

S16.1. Secondary purveyors must describe who (ie, the researcher) obtained the normative data they report.

S16.2. Secondary purveyors must describe in their reports where the normative data were obtained.

S16.3. Secondary purveyors must describe in their reports the sample studied to obtain the normative data.

S16.3.1. Secondary purveyors must describe in their reports how the sample used to obtain the normative data was selected.

S16.3.2. Secondary purveyors must specify in their reports the number of subjects studied to obtain the normative data.

S16.3.3. Secondary purveyors must explain in their reports how the sample used to obtain the normative data is characteristic of the population for whom the measurement is intended to be used.

S16.3.4. Secondary purveyors must describe in their reports relevant clinical characteristics of the sample used to obtain the normative data. These descriptions should include reports of the central tendencies, variabilities, and distributions of the data on relevant clinical, demographic, and anthropometric (physical) characteristics.

S16.4. Secondary purveyors must describe in their reports the persons who took the measurements used to obtain the normative data (ie, those who were in the role of test users). Descriptions of test users’ qualifications, competencies, and experiences with the test should be included. Any special information or training given to test users prior to their obtaining the
measurements in the study should be described by the secondary purveyor.

S16.5. Secondary purveyors must describe in their reports the methods and research designs used to obtain the normative data. Normative data should be obtained using the same measurement procedures that are described in the report. If other versions of the test were used to obtain the normative data, or if other scales were used, there must be a discussion of how the normative data relate to the data that can be obtained using the test described in the report.

S16.6. Secondary purveyors must supply in their reports a complete discussion of limitations in the use of the normative data they report. The discussion may include, but should not be limited to, considerations of whether the normative data relate to a particular local area, facility, ethnic group, age group, or gender.

S16.7. Secondary purveyors must supply in their reports details on any data transformations (e.g., any standardization or normalization procedures) used in obtaining and preparing the normative data they are reporting.

S16.8. Secondary purveyors who report normative data for measurements that are based on interval or ratio scales should present as part of the normative data standard scores or percentiles with accompanying measures of central tendency and variability. Data for clinically meaningful subgroups should be similarly reported.

S16.9. Secondary purveyors who report normative data for measurements that are based on ordinal or nominal scales should present the normative data in the form of the proportion of persons in the population who can be expected to belong to each group and subgroup. Data for clinically meaningful subgroups should be similarly reported.

S17. Advocacy by secondary purveyors for the use of a measurement must not exceed a level that can be supported by the research of the secondary purveyor or by other published data.
Standards for Tertiary Test Purveyors (indicated with a T)

The Standards in this section describe requirements for tertiary purveyors of tests. The following is the definition of a tertiary purveyor.

Tertiary purveyor: any person who teaches or prepares instructional material that describes specific tests or specific uses of measurements; this definition includes, but is not limited to, persons teaching in academic institutions, clinical educators, and continuing educators who are not acting in the role of primary or secondary purveyors (see purveyor, primary purveyor, and secondary purveyor)

Organization of the Standards for Tertiary Purveyors: Tertiary purveyors have two primary obligations: to understand tests and measurements in general and to have specific knowledge about the tests they discuss. The first part of these Standards describes general knowledge that a tertiary purveyor should have, and many of the subsequent Standards describe what information a tertiary purveyor should supply for each test the tertiary purveyor discusses. Tertiary purveyors, because they interact with potential test users, must be prepared to provide additional information to these potential users upon request. Some of the Standards describe the type of information that a tertiary purveyor should be prepared to supply during discussions.

T1. Persons should not become tertiary purveyors unless they are prepared to adhere to the Standards and unless they understand the requirements for primary and secondary purveyors. Persons should also not become tertiary purveyors unless they understand the requirements for test users and are willing to assist potential test users in complying with those Standards.

T2. Tertiary purveyors must have a basic knowledge of the theory and principles of tests and measurements.

T2.1. Tertiary purveyors must understand what constitutes a measurement, what constitutes a test, and the role of instruments in obtaining measurements.

T2.2. Tertiary purveyors must understand the differences between clinical opinions (impressions) that are not based on valid measurements and inferences that are based on the use of valid measurements.

T2.3. Tertiary purveyors must understand what constitutes an operational definition and the importance of using operational definitions.

T2.4. Tertiary purveyors must understand the different levels of measurement (ie, nominal, ordinal, interval, and ratio) and the mathematical operations that are appropriate for each level.

T2.5. Tertiary purveyors must understand types of reliability and validity and how these qualities relate to clinical decisions and other uses of measurements.

T2.6. Tertiary purveyors must understand the methods used to assess reliability and validity (eg, statistics and research designs).

T2.7. Tertiary purveyors must understand the relationship between reliability and validity and the differences between the two qualities.

T2.8. Tertiary purveyors must understand what constitutes meaningful normative data and how such data can be used.

T2.9. Tertiary purveyors must understand the differences between objective measurements and subjective measurements and the implications of using each type of measurement.

T2.10. Tertiary purveyors must understand the meaning and use of the terms "false negatives," "false positives," "true negatives," "true positives," "predictive value of a measurement," "specificity of a test," and "sensitivity of a test."

T2.11. Tertiary purveyors must understand the importance of knowing the technical specifications of instruments.

T2.12. Tertiary purveyors must understand the importance of calibrating instruments.

T2.13. Tertiary purveyors must understand the methods and effects of normalizing or standardizing measurements.

T2.14. Tertiary purveyors must understand the meaning and implications of reactivity to tests.

T3. Tertiary purveyors should promulgate the Standards for Test Users and, in their teaching, should provide potential test users with the necessary tools and information so that these potential test users can adhere to user standards.

T4. Tertiary purveyors, when discussing a test, must provide descriptions of the theoretical bases for the test and must discuss evidence relating to construct and content validity.
T5. Tertiary purveyors must provide all relevant operational definitions during their discussions of a test.

T5.1. Tertiary purveyors must provide operational definitions for attributes that the test measures.

T5.2. Tertiary purveyors must provide operational definitions for terms used to describe the population for whom the test is intended.

T5.3. Tertiary purveyors must provide operational definitions for terms used to describe potential test users.

T5.4. Tertiary purveyors must provide operational definitions for terms used to describe components of the test or test instrument.

T5.5. Tertiary purveyors must provide operational definitions for any terms created by purveyors.

T5.6. Tertiary purveyors must provide operational definitions for any terms they use in a noncustomary manner.

T5.7. Tertiary purveyors must provide operational definitions for any terms they modified or created.

T6. Tertiary purveyors, during discussions of a test, must provide a description of the population for which the test is designed. Descriptions of subjects for whom the test should not be used and descriptions of subjects for whom the test should be used with caution should be included.

T7. Tertiary purveyors have an obligation to review critically what is known about the reliability and validity of tests that they discuss, including how statistics were used to assess reliability and validity. Tertiary purveyors must also be prepared to answer questions of potential test users regarding reliability and validity studies and statistics used in these studies.

T8. Tertiary purveyors must provide all information that is available when they convey or discuss normative data.

T8.1. Tertiary purveyors must describe the methods used to obtain the sample that was used to obtain the normative data. The generalizability of the normative data must be characterized relative to the sampling method.

T8.2. Tertiary purveyors must describe the sample studied (e.g., the number of subjects and the distributions of relevant clinical, demographic, and anthropometric [physical] characteristics). How this group is characteristic of the population for whom the test is intended must also be discussed.

T8.3. Tertiary purveyors must discuss limitations in the normative data. This discussion may include, but should not be limited to, considerations of whether the data relate to one local area, facility, ethnic group, age group, or gender.

T8.4. Tertiary purveyors must discuss details on any data transformations used and whether any standardization or normalization procedures were used in generating the normative data.

T8.5. Tertiary purveyors who discuss measurements that are interval or ratio scaled must provide standard scores or percentiles with measurements of central tendency and variability, if these data are available. Data for meaningful subgroups should be similarly reported. If these data are lacking, the tertiary purveyor should discuss the limitations in the use of the normative data.

T8.6. Tertiary purveyors who discuss measurements that are ordinal or nominal scaled or who describe classifications must provide normative data in terms of the proportion of persons in the population that can be expected to belong to each group, if this information is available. Data for meaningful subgroups should be similarly reported. If these data are lacking, the tertiary purveyor should discuss the limitations in the use of the normative data.

T9. Tertiary purveyors, in discussing a specific test, must provide descriptions of the qualifications and competencies needed by the test user to administer that test.

T10. Tertiary purveyors, when discussing a specific test, should provide a brief account of the development of the test.

T11. Tertiary purveyors, in discussing a specific test, must provide descriptions of the test and instruments associated with the test.

T11.1 Tertiary purveyors must discuss available documentation of relevant technical information regarding performance characteristics of any machines, recording devices, transducers, computer interfaces, and similar instruments. The tertiary purveyor should identify the source of this documentation. If documentation is not available, the tertiary purveyor must discuss the implications and limitations of using such instruments.

T11.2. Tertiary purveyors must describe how instruments used in the test manipulate or process data.
information in order to obtain the measurements, if this information is available. Tertiary purveyors should identify the source of this information. If this information is not available, the tertiary purveyor must discuss the implications and limitations of using such instruments.

T12. Tertiary purveyors must provide instructions for administering all tests that they teach to potential test users. These instructions must include descriptions of the sources for test manuals as well as all equipment and activities needed for obtaining, recording, and interpreting the measurements.

T12.1. Tertiary purveyors must provide guidelines for what information and instructions are to be given to the person being tested. Information about the test should be provided that will allow the potential test user to answer questions about the test and related subjects.

T12.2. Tertiary purveyors must describe the physical settings in which the test should be given and the possible effects of conducting the test in other settings.

T12.3. Tertiary purveyors must describe conditions, behaviors of persons taking the test, or other factors that could make the validity of the measurements questionable.

T12.4. Tertiary purveyors must describe how the test user must manipulate or process information in order to obtain the desired measurements.

T12.5. Tertiary purveyors must provide instructions to potential test users for use of any instruments required to obtain the desired measurements. These instructions, where appropriate, must include machine settings and any other user-selected options. Descriptions of the effects of all options on the measurements and the consequences of selecting the incorrect options should be included.

T12.6. Tertiary purveyors who discuss tests that involve the use of instruments must describe how the instruments are calibrated. A means of testing calibration must be described. If calibration is needed, the tertiary purveyor must provide instructions regarding a course of action to be taken.

T12.7. Tertiary purveyors must describe variations in the test procedures that are available to the test user. Descriptions of variations that are known not to impair the quality of the measurements and descriptions of variations that are known to lead to measurements of questionable validity must be included.

T12.8. Tertiary purveyors must provide background information so that potential test users will have the knowledge to obtain any derived measurements or categorization necessary for interpretation of the measurements.

T13. Tertiary purveyors must provide warnings regarding misuse of the measurements they discuss. Common errors in interpretation of the measurements must be described. If research or the tertiary purveyor’s experience indicates that errors in interpretation of test data can occur, then these errors should be described.

T14. Tertiary purveyors must discuss the implications of reactivity when discussing a test.

T14.1. Tertiary purveyors must discuss the degree to which administration of the test affects the measurement obtained from that test or any subsequent tests.

T14.2. Tertiary purveyors must discuss the degree to which administration of the test may cause a change in the person taking the test. Discussions of side effects, aftereffects, fatigue, learning, and so forth may be included.

T15. Tertiary purveyors must include in their discussions of a test descriptions of all special groups for whom the test is contraindicated or known to lead to measurements of questionable validity.

T16. Tertiary purveyors, when discussing measurements used to classify persons into groups based on the presence or absence of a diagnostic finding (eg, use of cut scores or tests to determine a positive or negative finding), must discuss the limitations of these measurements.

T16.1. Tertiary purveyors must report the percentages of false positives and false negatives for the measurements they discuss. If this information is not available, tertiary purveyors must discuss the limitations of using these measurements.

T16.2. Tertiary purveyors must report the sensitivity of the tests they discuss. If this information is not available, tertiary purveyors must discuss the limitations of using these tests.

T16.3. Tertiary purveyors must report the specificity of the tests they discuss. If this information is not available, tertiary purveyors must discuss the limitations of using these tests.
T16.4. Tertiary purveyors must report the predictive values of positive and negative findings for the measurements they discuss. If this information is not available, tertiary purveyors must discuss the limitations of using these measurements.

T17. Tertiary purveyors, when they discuss a test, must identify any way in which their versions of the test differ from published versions of the test. Tertiary purveyors must also discuss how these variations can affect the measurement and the uses of the measurement. A tertiary purveyor who modifies a test becomes a primary purveyor and must meet the Standards specified for primary purveyors.

T18. Tertiary purveyors must provide information that will enable potential test users to understand the limitations of tests that do not meet the Standards. Tertiary purveyors must make potential test users aware of the limited justifiable inferences that can be made from tests that do not meet the Standards.

T19. Tertiary purveyors must discuss with potential test users issues related to the interpretation of the measurements they discuss. Tertiary purveyors must warn potential test users of common errors that the purveyors know occur in clinical practice. Strategies for avoiding these errors should be discussed.

T20. Tertiary purveyors must discuss with potential test users how results of tests must be reported. Tertiary purveyors must discuss what information is essential in reports.

T21. Tertiary purveyors must discuss with potential test users the difference between clinical opinions and interpretations that are based solely on valid measurements. The tertiary purveyor must also provide examples of how clinical opinions may be differentiated from test findings in clinical reports and other communications.

T22. Tertiary purveyors must assist potential test users in understanding the role of specific measurements in the clinical decision-making process. Tertiary purveyors must characterize whether existing research justifies conclusions based on single tests or whether clinical decisions should be the result of the synthesis of multiple measurements.
Standards for Test Users
(Indicated with a U)

The Standards in this section describe requirements for test users. The following is the definition of a test user.

**Test user:** one who chooses tests, interprets test scores, or makes decisions based on test scores (this definition is from Standards for Educational and Psychological Tests. American Psychological Association, Washington, DC, 1974, page 1)

**Organization of the Standards for Test Users:** Four basic types of Standards are found in the Standards for Test Users. The Standards listed first detail the general knowledge that a test user must have. The majority of the Standards in this section deal with specific requirements that a user should consider when performing specific tests. These Standards include issues relating to the choice of tests, the performance of testing, observing the rights of test takers, and the use of obtained measurements. The last two Standards, U44 and U45, describe the requirements test users should observe in interpreting and reporting test results.

**U1.** Persons should not become test users unless they are prepared to adhere to the Standards and understand the requirements for test purveyors.

**U2.** Test users must have a basic understanding of local, state, and federal laws governing the use of tests in their practice settings.

**U3.** Test users must have a basic knowledge of the theory and principles of tests and measurements.

**U3.1.** Test users must understand what constitutes a measurement, what constitutes a test, and the role of instruments in obtaining measurements.

**U3.2.** Test users must understand the differences between clinical opinions (impressions) that are not based on valid measurements and inferences that are based on the use of valid measurements.

**U3.3.** Test users must understand what constitutes an operational definition and the importance of using operational definitions.

**U3.4.** Test users must understand the different levels of measurement (ie, nominal, ordinal, interval, and ratio) and the mathematical operations that are appropriate for each level.

**U3.5.** Test users must understand types of validity and how these types of validity relate to the use of measurements.

**U3.6.** Test users must understand types of reliability and validity and how these qualities relate to clinical decisions and other uses of measurements.

**U3.7.** Test users must have a basic understanding of the methods used to assess reliability and validity (eg, statistics and research designs).

**U3.8.** Test users must understand the relationship between reliability and validity and the differences between the two qualities.

**U3.9.** Test users must understand what constitutes meaningful normative data and how such data can be used.

**U3.10.** Test users must understand the differences between objective measurements and subjective measurements and the implications of using each type of measurement.

**U3.11.** Test users must understand the meaning and use of the terms “false negatives,” “false positives,” “true negatives,” “true positives,” “predictive value of a measurement,” “specificity of a test,” and “sensitivity of a test.”

**U3.12.** Test users must understand the importance of knowing the technical specifications of instruments.

**U3.13.** Test users must understand the importance of calibrating instruments.

**U3.14.** Test users must have a basic understanding of the methods and effects of normalizing or standardizing measurements.

**U3.15.** Test users must understand the meaning and implications of reactivity to tests.

**U4.** Test users must have background knowledge in basic, applied, and clinical sciences related to the selection, administration, and interpretation of each test they use.

**U5.** Test users must understand the theoretical bases (construct and content validity) for the tests they use, and they must have knowledge about the attribute (characteristic) being measured.

**U6.** Test users must be familiar with the development of tests that they use and the test settings in which those tests have been developed and used.

**U7.** Test users must understand how a test they are using relates to similar tests or previous versions of the same test.
US. Test users must be able to justify the selection of tests they use. Test users must also be prepared to supply logical arguments to justify the rejection of tests they choose not to use.

U8.1. Test users must consider the safety of subjects in selecting tests and should consider the benefits to be obtained from a test in view of potential risks to the subject.

U8.2. Test users should consider the practicality of the test (e.g., personnel, time, equipment, cost of administration, and impact on the person taking the test) in selecting tests and in planning examination procedures.

U9. Test users must be able to identify their sources of information regarding tests they use. Test users must be able to specify where they obtained information (e.g., rationale and directions) for selecting and conducting a test.

U9.1. Test users should not cite a test manual as a source of information unless they have personally examined a complete copy of the test manual. Test users should not conduct tests unless they have examined all relevant sections of a complete copy of the test manual.

U10. Test users must understand all operational definitions related to tests they use.

U10.1. Test users must understand the operational definitions for attributes that the test measures.

U10.2. Test users must understand the operational definitions for terms used to describe the population for whom the test is intended.

U10.3. Test users must understand the operational definitions for terms used to describe potential test users.

U10.4. Test users must understand the operational definitions for terms used to describe components of the test or test instruments.

U10.5. Test users must understand the operational definitions for any terms created by purveyors of the test.

U10.6. Test users must be able to identify and understand the operational definitions for any terms used in a noncustomary manner.

U11. Test users must be able to describe the population for whom the test was designed. Test users must be able to relate this description to the persons they are testing.

U12. Test users must be able to determine before they use a test whether they have the ability to administer that test. This determination should be based on an understanding of the test user's own skills and knowledge (competency) as compared with the competencies described by the test purveyor.

U12.1. Test users must be able to describe the potential consequences of administering a test that they do not have the skills or knowledge to administer.

U12.2. Test users who have doubts about their ability to administer a test should report this information when they report test results (e.g., their reservations about the quality of their measurements should be discussed).

U13. Test users must follow instructions provided by purveyors for all tests they administer.

U13.1. Test users must understand instructions for administering all tests that they use. Test users must be able to describe all of the equipment and activities needed for obtaining, recording, and interpreting the measurements. Test users must be able to identify the source of the instructions.

U13.2. Test users who deviate from accepted directions for obtaining a measurement should not use published data or documentation relative to reliability and validity to justify their use of the measurement.

U14. Test users must know what information and instructions are to be given to the person being tested. Test users should be able to answer questions about the test and related subjects.

U14.1. Test users who do not give the purveyor's specified instructions to persons being tested, or test users who are unable to give these instructions, should not use published data or documentation relative to reliability and validity to justify their use of the measurements.

U15. Test users must know the physical settings in which the test should be given and the possible effects of conducting the test in other settings.

U16. Test users must be able to identify any conditions or behaviors in the person being tested that may compromise the reliability or validity of their measurements (e.g., if a modified position must be used in manual muscle testing because of a deformity). Test users who observe such conditions or behaviors should note these observations in their reports of any resultant measurements. Test users who be-
lieve that the effect on their measurements could be significant should include a discussion of the implications of these observations in their reports.

U17. Test users must have a basic understanding of the instruments they use as part of a test.

U17.1. Test users must know relevant technical information regarding performance characteristics of any machines, recording devices, transducers, computer interfaces, and similar instruments they use. Test users should be able to identify the source of this information. If this information is not available, the test user must be able to discuss the implications and limitations of using such instruments.

U17.2. Test users must be able to describe how instruments they use manipulate or process information in order to obtain measurements. Test users should identify the source of this information. If this information is not available, the test user must be able to discuss the implications and limitations of using such instruments.

U18. Test users must know how to use any instruments required to obtain the desired measurements. This Standard includes, where appropriate, the test user knowing how to choose machine settings and other user-selected options. Test users must be able to discuss the effects of all options on their measurements and the consequences of selecting the incorrect options.

U19. Test users must be able to describe how instruments they use for a test are calibrated, including the means of testing calibration. Test users must know the course of action to be taken when calibration is needed.

U20. Test users, for all the tests they use, should be able to describe variations in the test procedures that are available. Test users must be able to describe variations that are known not to impair the quality of the measurements and those variations that are known to lead to measurements of questionable validity.

U21. Test users who deviate from accepted directions for obtaining a measurement should not use published data or documentation relative to reliability and validity to justify their use of the measurement.

U21.1. Test users who administer tests in settings other than those recommended by the purveyor should not use published data or documentation relative to reliability and validity to justify their use of the measurement.

U22. Test users have a responsibility to suggest further testing when they have serious concerns about the quality of the measurements they obtain or when they believe that other tests or other personnel can be used to obtain better measurements.

U23. Test users who are required to derive or transform measurements must have sufficient training and knowledge to derive or transform those measurements. Test users must have the background information and skills needed to derive measurements or make categorizations necessary for interpretation of their measurements (eg, how to normalize or standardize a score or how to classify a measurement).

U24. Test users must be aware of any normative data for the measurements they are obtaining (see Standard U44.3 for guidelines on using normative data to interpret measurements; see Standard U45.10 for guidelines on reporting measurements related to normative data). Test users should be able to evaluate critically normative data and use the data for clinical decision making.

U25. Test users must make every effort to control the environment (test setting) in which they test in order to maintain consistent conditions between tests. These efforts are needed to ensure that the validity and reliability of a measurement are not compromised.

U26. Test users must make every effort when personal information is being obtained to control the environment (test setting) in which they administer tests in order to preserve the privacy of the person taking the test.

U27. Test users must be able to discuss common errors in the interpretation of the measurements they use.

U28. Test users must make every effort to minimize the effects of reactivity associated with the tests they use.

U29. Test users should report to the purveyor of the test any problems regarding a test or any associated instruments.

U30. Test users should communicate with other test users and purveyors regarding their experiences with tests.

U31. Test users must avoid giving persons prior knowledge about the nature of a test when such knowledge is known to compromise the validity of the measurements.
U32. Test users are responsible for maintaining confidentiality of test results. Confidentiality of results should be in accordance with standard practices in the institution or community in which the test user obtains the measurements. Results should not be shared with any persons (or organizations) who are known to be unwilling to respect the right of confidentiality of the person who was tested.

U33. Test users should not share results of tests with persons (or organizations) who are likely to misuse that information.

U34. Test users must respect the rights of persons whom they test.

U34.1. Test users must respect the right of persons to refuse to be tested. Test users must allow persons to discontinue participation in any test at any time without recrimination or prejudice against that person.

U34.2. Test users must inform persons whom they test of potential risks and benefits that persons may experience as a result of taking the test.

U34.3. Test users must respect the right of persons being tested to know the results of tests, the interpretations of those test results, and with whom the test results will be shared. The right of the person to know the results of tests does not imply that all test users must personally supply this information. In some cases, test results may be supplied by the professional who originated a referral or who is coordinating treatment.

U34.4. Test users who fail to adhere to the Standards and who use tests inappropriately, especially in terms of drawing unwarranted conclusions from results, violate the rights of persons being tested.

U34.4.1. Test users who misrepresent their clinical opinions as being based on test results when evidence for such opinions is not found in the research literature violate the rights of persons taking tests. (For example: A test user may use a battery of tests to determine the ability of a patient with low back pain to function in an industrial environment. In this hypothetical example, the test battery yields a measurement that is supposed to predict the type of work that the patient may do safely. There is, in this example, evidence for the validity of this inference. However, based on the test user's observations, the test user concludes that the patient is malingering. This is the test user's clinical opinion; it is not based on the validated use of the measurement. The test user does not violate the rights of the person taking the test by having or presenting clinical opinions, but would violate the person's rights by contending that the measurement could be used to infer malingering.)

U35. Test users must maintain records in such a manner that information about tests and measurements is accurate and is not likely to be distorted or lost. Abbreviations used in communications should be limited to those that appear in established references.

U36. Test users have a responsibility to report inappropriate test use to proper authorities.

U36.1. Test users who know that a person's rights are not being observed during testing must make every effort to change that situation.

U37. Test users should select tests based on what is best for the person being tested. Test selection based on considerations of personal benefit to the test user, test purveyor, or the referring practitioner is inappropriate.

U38. Test users, in clinical practice, should avoid the use of tests that were designed solely for research purposes. Such tests, when they are used in the clinical setting, should be identified in all reports as research tests that have not necessarily been shown to be reliable or valid in clinical use.

U39. Test users should not assign persons to conduct tests unless they know that such persons are qualified to conduct the tests.

U40. Test users should not make promotional claims for their testing procedures that are not supported by research literature.

U40.1. Test users are responsible for the critical evaluation of all claims of test purveyors and should not merely repeat the claims of purveyors without critical evaluation of these claims.

U41. Test users should assist in the development and refinement of testing procedures by sharing their knowledge of tests and assisting in the collection of data where appropriate.

U42. Test users have a responsibility to periodically review the test procedures they and their colleagues use in their institutions (practice settings) to ensure that appropriate use of measurements is being
made and that the rights of persons tested are being observed.

U42.1. Test users, as part of their periodic review of test procedures, should examine whether the normative data they are using appear to relate to their clinical setting.

U42.2. Test users, as part of their periodic review of test procedures, should attempt to estimate the reliability of measurements in their practice settings. All forms of reliability relevant to the practice settings should be assessed.

U43. Test users who use tests that do not meet the Standards should be aware that these tests do not meet the Standards. Test users, therefore, should interpret results of these tests with caution and share these reservations with all persons who receive test results.

U44. Test users must follow the basic rules and principles of measurement when they interpret results of tests they use. (The following Standards provide guidelines for interpreting measurements. These Standards are not meant to supersede or in any way modify the requirements specified elsewhere in the Standards for Test Users.)

U44.1. Test users must limit their interpretations of measurements to the inferences for which those measurements have been shown to be valid.

U44.2. Test users must consider the error associated with their measurements when they interpret their test results. Reliability and validity estimates should be considered when the test user makes interpretations of measurements. (For example: Reliability studies have indicated that a measurement varies as much as 10% between repeated tests. Therefore, a change of less than 10% on that measurement may be due, at least in part, to measurement error. Test users who note changes the second time they take measurements should consider, before they make interpretations, that the change may not reflect real change, but may be due solely to measurement error.)

U44.3. Test users must consider whether normative data are available for the measurements they interpret. Test users must consider the sources of the normative data and how applicable these data are to the measurements they are interpreting.

U44.3.1. Test users should use all available information when using normative data for interpretations of measurements.

U44.3.1.1. Test users using normative data should interpret any measurement that is interval or ratio scaled in terms of how that measurement relates to measures of central tendency, measures of variability, and percentiles.

U44.3.1.2. Test users using normative data should interpret any measurement that is nominal or ordinal scaled in terms of the proportion of persons in the population that can be expected to belong to the same classification.

U44.4. Test users must consider the limitations of their measurements when they classify persons into diagnostic groups based on the presence or absence of a finding (eg, use of cut scores or tests to determine a positive or negative finding). Test users should use all available data in making their interpretations.

U44.4.1. Test users must consider the percentages of false positives and false negatives for a diagnostic test when interpreting measurements. If this information is not available, test users should understand the limitations of making interpretations based on their measurements.

U44.4.2. Test users must consider the sensitivity of the diagnostic test they are using when they interpret their measurements. If this information is not available, test users should understand the limitations of making interpretations based on their measurements.

U44.4.3. Test users must consider the specificity of the diagnostic test they are using when they interpret their measurements. If this information is not available, test users should understand the limitations of making interpretations based on their measurements.

U44.4.4. Test users must consider the predictive values of positive and negative findings when they interpret their measurements obtained with a diagnostic test. If this information is not available, test users should understand the limitations of making interpretations based on their measurements.
U44.5. Test users must avoid overinterpreting the results of their tests. Test users are responsible for understanding both the certainty and the uncertainty with which they can make judgments based on their measurements.

U44.6. Test users must consider whether changes (eg, attributable to development or learning) in the person being tested may alter performance on subsequent tests. Test users, when appropriate, should discuss in their reports of test results the possibility of change in the future. Test users should not imply that a test result represents an immutable state when there is reason to believe that the test result may differ if the test is repeated at some future time.

U44.7. Test users must consider the conditions under which they conduct tests and the extent to which results are generalizable to other test situations (eg, testing in other places or at other times).

U44.8. Test users must identify whether their interpretations are based on the results of multiple measurements obtained with the same test or on the results of a single measurement.

U44.9. Test users must identify whether any of their interpretations are not supported by research evidence of validity. Such interpretations must be clearly identified as being based on the test user's personal opinion.

U45. Test users reporting the results of tests must supply adequate information so that these results can be understood. (The following Standards provide guidelines for reporting about measurements. These Standards are not meant to supersede or in any way modify the requirements specified elsewhere in the Standards for Test Users.)

U45.1. Test users should specify, when more than one form of a test exists, the specific form of the test used when they report their results.

U45.2. Test users should report measurements in the form specified by the purveyor's instructions. Test users should justify any deviations from standard methods of reporting.

U45.3. Test users should use only the terms that are defined in test manuals or in other supporting literature when they discuss tests or measurements. Descriptive terms that are not defined should be avoided, because such terms may encourage inappropriate interpretation of results.

U45.4. Test users, in reporting test results, should use terms in a customary manner or describe how terms are being used differently. Test users should justify deviations from commonly accepted uses of terms in their reports.

U45.5. Test users must consider estimates of reliability and validity when reporting test results. Test users should report estimates of the errors associated with a measurement when they report test results. (For example: Reliability studies have indicated that a measurement varies as much as 10% between repeated measurements. Therefore, a change of less than 10% may be due, at least in part, to measurement error. Test users who note changes the second time they take measurements, in reporting such measurements, should also report that the change in the measurement may not reflect real change. The change may be solely due to measurement error. A report of the reliability estimate or standard error, in this case, would be useful in the test user's report.)

U45.6. Test users should include warnings about common misinterpretations of their measurements in reports of their measurements.

U45.7. Test users should report any significant effects of reactivity when they report the results of their tests.

U45.8. Test users who use a variation of a test must indicate, when they report test results, that a variation was used. The test users must note whether they believe that the variation may have affected the quality of their measurements. Test users who believe the variation had a significant effect on the measurements should discuss this belief in all reports of test results.

U45.9. Test users should report any aspect of the test that may cast doubt on test results (eg, ways in which the person tested differed from the population for which the test was designed or any observation the test user made during testing).

U45.10. Test users, in reports of test results, should relate their measurements to normative data, if available. Test users should report the source of the normative data they use and, if necessary, discuss how applicable the data are to the measurements they are reporting (see Standard U44.3 for guidelines on using normative data to interpret measurements).

U45.10.1. Test users using normative data, when they report test results, should report all information necessary to understand
the test user's interpretation of the measurements.

U45.10.1.1. Test users using normative data for measurements that are interval or ratio scaled should report their test results in terms of how the measurements relate to measures of central tendency, measures of variability, and percentiles.

U45.10.1.2. Test users who report classifications in their test results should also report the proportion of persons in the population who can be expected to belong to that classification. The test user, if requested, should be able to cite the source of the data used to determine the proportions.

U45.11. Test users reporting the results of their tests should indicate whether any data were transformed (normalized or standardized). Test users, in their reports, should justify the use of transformations, if this is not customary practice.

U45.12. Test users who base their interpretations of test results on the mean of multiple measurements should note this fact in their reports of test results. Test users should justify the use of the mean of multiple measurements in clinical reports, if this is not customary practice.

U45.13. Test users who base their interpretations on a single measurement chosen from a group of measurements (e.g., the best of three trials) should note this fact when they report test results. Test users, in their reports, should justify the use of the single measurement and the criteria used to select the measurement, if this is not customary practice.

U45.14. Test users who base their interpretations on the results of a variety of tests should note this fact when they discuss their measurements. Test users should justify their selection of the tests in reporting test results.

U45.15. Test users should note in their reports of test results the specific criteria they use for clinical decisions. When a specific measurement (e.g., cut score) is used for a clinical decision, the test user, in all reports, should justify the use of that specific measurement.
Standards for Ensuring Integrity in Measurement Research (Indicated with an R)\textsuperscript{1}

R1. Physical therapists who conduct measurement research should maximize the integrity of their work by following the guidelines set forth in the Standards for Tests and Measurements sections on primary and secondary purveyors.\textsuperscript{2}

R2. Physical therapists must ensure that subjects in measurement studies are volunteers and that no coercion or deception was used to entice subjects to volunteer. Participation of each volunteer should be based on the subject's (or the subject's legally authorized representative's) understanding of the nature of the study and its expected risks and benefits.

R2.1. Physical therapists who conduct measurement research must obtain, in writing, informed consent from subjects or the subjects' legal representatives before the subjects participate in studies. The form of the consent must be in accord with appropriate laws, regulations, and institutional requirements.

R2.2. Physical therapists who conduct measurement research must assure their subjects or the subjects' legal representatives, in writing, of the subjects' right to withdraw consent and discontinue participation in the measurement study. Such withdrawal should not result in any prejudice against or negative impact upon the subject.

R2.3. Physical therapists who obtain informed consent as part of measurement research must inform their subjects or the subjects' legal representatives, in writing, of the extent to which confidentiality will be maintained.

R3. Physical therapists must ensure that information about subjects obtained during measurement research is recorded, stored, and reported in ways that protect the subjects' right to confidentiality.

R3.1. Physical therapists who conduct measurement research, before they use any information about the subjects, must inform their subjects or the subjects' legal representatives about this planned action. The subjects or the subjects' legally authorized representatives must authorize, in writing, the release of this information. This information includes any data or recorded images of the subjects.

R4. Physical therapists who conduct measurement research are expected to ensure the privacy of subjects during the course of their measurement research. Therapists, if more than one subject must be present during a test session, should ensure that each subject has the maximum possible privacy.

R5. Physical therapists who conduct measurement research must minimize the risk of physical, psychological, or social harm to their subjects.

R6. Physical therapists who conduct measurement research must be guided at all times by a concern for the physical, psychological, and social well-being of their subjects.

R7. Physical therapists who use patients as subjects during measurement research must comply with the applicable laws regulating the practice of physical therapy in the jurisdiction in which the study is taking place.

R8. Physical therapists must make every effort to comply with the requirements governing the approval and conduct of measurement research within the institutional or organizational setting in which they conduct research. If the setting has no requirements governing the approval of proposals for measurement studies, physical therapists should assist in developing and implementing such requirements.

R9. Physical therapists who seek institutional approval and external funding for measurement research must provide accurate information to institutional review boards, funding agencies, and other relevant groups.

R10. Physical therapists who conduct measurement research that has been approved by an institutional review board or a funding agency are obligated to adhere to the approved protocol and to deviate from that protocol only in accordance with the policies of the board or agency that granted approval.

R11. Physical therapists who conduct measurement research must ensure that reports of their studies are accurate and represent only the work that was done in the study.

R12. Physical therapists must conduct measurement research in such a way that the prospect of financial gain or past financial assistance to the investigators or their institutions has no influence on the results or the manner in which the results are reported.

\textsuperscript{1}These Standards are adapted from the document Integrity in Physical Therapy Research, which was approved by the APTA Board of Directors in March 1985 and modified in November 1987.

\textsuperscript{2}The Standards for Ensuring Integrity in Measurement Research describe specific requirements that should be met by researchers. Researchers may be primary or secondary test purveyors and will, almost always, be test users; therefore, researchers must comply with all relevant Standards described in those sections.
R12.1. Physical therapists must report all sources of financial support for a study when the results of their study are published.

R13. Physical therapists must make every effort to share information about their measurement studies and findings in an appropriate manner.

R13.1. Physical therapists should submit formal research reports to journals in which the manuscript will be subjected to meaningful peer review.

R13.2. Physical therapists should honor the requests of professional colleagues for access to their measurement research data.

R14. Physical therapists must have a thorough knowledge of measurement theory and pertinent professional and scientific literature before they conduct measurement studies. Measurement studies must be carried out with due consideration given to this body of knowledge.

R15. Physical therapists who conduct measurement research are obligated to know their personal limitations and to know when to seek consultation and peer review of their research plans before they begin data collection.

R16. Physical therapists, in written reports of measurement studies, must ensure that all references are correct and that citations are used appropriately. References must directly support information in the sentence in which the references are cited. If the references supply indirect support for the statement, this support should be indicated.

R17. Physical therapists must describe, in written reports of measurement studies, all relevant information in adequate detail so that readers may replicate the study.

R18. Physical therapists must describe, in written reports of measurement studies, steps that were taken to comply with governmental and institutional regulations governing research in the setting in which the study was carried out. The report should include descriptions of the steps that were taken to ensure subjects' rights.

R19. Physical therapists who are participating in measurement studies must dissociate themselves from any activities that are unethical or unlawful. Physical therapists must make every effort to take corrective action when they encounter unethical, unlawful, or incompetent acts in the conduct of measurement research. In addition, they are obligated to report any unethical, unlawful, or incompetent acts of any person to the appropriate authorities.

R20. Physical therapists must assist the profession in documenting the integrity of measurement research. Physical therapists called upon to assist in inquiries about any measurement research should cooperate. There should be no recriminations against any person called upon to serve as an investigator. A person should not investigate any other person's work unless that person possesses the necessary knowledge and skills to evaluate objectively all available data.