Measurement in Physical Therapy

On the Rules for Assigning Numerals to Observations

EUGENE MICHELS

This paper discusses certain issues of theory, concept, definition, and method in measurement that are of concern to physical therapy. The topics discussed include the place of measurement in science, definitions of measurement, direct and indirect measurement, the logical requirements of measurement, scales of measurement, precision and accuracy, and reliability and validity. Comments pertinent to physical therapy are included in the discussion, and the issues are summarized by a look at the complex problems of using EMG for measurement of muscle activity.

Key Words: *Measurement, Physical therapy, Research.*

One of our colleagues commented recently on knowledge of measurement as a neglected aspect of physical therapy education. She emphasized the need for knowledge of formal measurement theory in planning research and in using tests appropriately in clinical situations.[1]

Going one step further, the literature in physical therapy contains no serious discussion of measurement. By "serious discussion" I mean critical review or logical analysis of issues of theory, concept, definition, and method in measurement. This article addresses some of the issues that are or should be of concern in physical therapy.

MEASUREMENT AS MEANS OR END

The recent comments by our colleague, referred to above, suggest that knowledge of formal measurement theory is a means to the end of planning successful research.[1] This view of the place of measurement in science finds support in comments made by others outside our field: "Matter is understood most competently with measurement and numbers."[2 (p53)] "Quantification is adopted because it is more adequate to the description of phenomena [and it] is favored by the desire of investigators to claim the prestige of science for their research."[3 (pp124–125)] Given the current state of measurement and research in physical therapy, we can readily subscribe to the view that measurement and knowledge of measurement theory are important and necessary means to the end of achieving our science.

But there are other views to consider: some that caution us about being overly zealous in quantifying everything, and others that make us ponder the place of measurement in our science.

First, the words of caution: "The trouble with the idea of measurement is its seeming clarity, its obviousness, its implicit claim to finality in any investigative discourse."[4 (p163)] "'Measurement' is one of those terms which has attained a social prestige. Apparently—all other things being equal—it is better to measure than not to measure."[5 (p83)] The latter view appears to be widely held among people doing research in physical therapy. That view is neither good nor bad unless it is accompanied by measuring for the sake of measuring. From a history of quantification in medical science we learn that measuring for the sake of measuring may be one of the early stages in the development of a science. "Perhaps such procedure may occur in any field. But it is to be expected especially in an area wherein demand for quantification arises before qualitative preparation has been adequate."[6 (p91)] Here lies our first glimpse of the possibility that measurement may not be only a means to the end of achieving a science.

Now consider the place of measurement in science: "Quantification in science ... must follow, not precede, adequate qualification. Measurement can be helpful only when the proper things have been found

Mr. Michels is Associate Executive Director for Research and Education, American Physical Therapy Association, 1156 15th St NW, Washington, DC 20005 (USA).

From a paper presented at the Section on Research, The Eugene Michels Researchers' Forum, 1982 Combined Sections Meeting, San Diego, CA, February 14, 1982.

This article was submitted February 22, 1982, and accepted July 13, 1982.

to measure. It derives from, feeds into, sharpens and clarifies, and discriminates between alternate qualitative descriptions and models; it cannot generate them."[7 (p204)] "If measurement ever leads to discovery or to confirmation [of theory], it does not do so in the most usual of all its applications.... To discover quantitative regularity one must normally know what regularity one is seeking and one's instruments must be designed accordingly.... *The road from scientific law to scientific measurement can rarely be traveled in the reverse direction.*"[8 (pp41,59–60)]

The latter ideas suggest that measurement may be an end to be achieved by a science. More than that, the ideas suggest that theory and concepts—not just measurement theory—are preliminary to measurement that is scientifically useful. The intellectual work on theory and concepts is what is meant by "qualitative preparation" and "adequate qualification." The idea that adequate qualification must precede quantification is a challenge to think more rigorously about what it is that our measurement operations are intended to measure. And in our thinking we would do well to keep before us the simple but profound statement that "quantities are *of* qualities."[9 (p207)]

DEFINITIONS OF MEASUREMENT

Serious consideration should also be given to what we mean by the term "measurement." Be assured that there is no universally agreed upon meaning of the term.

"In the case of measurement, first of all, it has not always been clear whether the term means an *operation* involving an observer and a more or less complex apparatus, or whether it means the *number* that emerges as the result of such an operation—whether, in other words, a measurement produces a result or an operation produces a measurement."[10 (p4)] Even the word "number" used in connection with measurement is subject to some dispute, and some authors take great care to distinguish between numerals and numbers.[11]

The definition of measurement that I have found most useful is this: measurement is the act of converting observations into data,[12] and includes classifying, counting, ranking, and quantifying. There are other brief definitions. Stevens, recognized widely as the leading advocate, and by some the original advocate, of four scales of measurement (nominal, ordinal, interval, and ratio), defined measurement as the assignment of numerals to objects or events according to rule—any rule.[13, 14] Some other definitions are as follows: Measurement is the process of assigning numerals to represent properties or qualities.[11 (p267)] "Measurement consists of rules for assigning numbers to objects to represent quantities of attributes."[15 (p2)] "Measurement is the assignment of particular mathematical characteristics to conceptual entities in such a way as to permit 1) an unambiguous mathematical description of every situation involving the entity and 2) the arrangement of all occurrences of it in a quasi-serial order . . . an order which determines, for any two occurrences, either that they are equivalent with respect to the property in question or that one is greater than the other."[10 (p5)]

Some definitions of measurement use the word "rule." A measurement rule is said to be "a systematic rule of procedure that permits one to identify each possible event that might occur in the given observational situation with one of a set of different categories or symbols."[16 (p4)] This systematic procedure is also referred to as the operational definition of the dimension of interest, the scale of measurement to be used, and the categories or units on that scale. Whether the procedure is determined by rule or by operational definition, there seems to be some wisdom worth considering in the statement, "A procedure of measurement not only determines an amount, but also fixes what it is an amount *of*.... What is measured and how we measure it are determined jointly."[9 (p177)]

Most definitions of measurement use or imply the use of the terms "numbers," "numerals," and "quantities." I avoid the use of these terms in my preferred definition because classifying, which is one form of measurement, neither requires nor implies the assignment of numerals or numbers to represent quantities. In fact, the assignment of numerals to categories on the nominal or ordinal scale—especially the ordinal scale—produces all kinds of mischief because people assume, probably without careful thinking, that such numerals represent numbers or amounts of whatever the scale is intended to measure.

Definitions of measurement have ultimately to do with the logical requirements of measurement and with the logic of the scales of measurement that one is willing to admit into the domain of measurement. Before turning to these logical matters, consider the kinds of measurement that others have identified and that have a place in any subsequent serious discussions of measurement in physical therapy.

KINDS OF MEASUREMENT

Campbell identified two kinds of measurement, fundamental and derived.[11] Ellis, considering only the measurement of quantity, extended Campbell's ideas by identifying direct measurement (Campbell's fundamental measurement) and indirect measurement, the latter including derived measurement and associative measurement.[14]

Fundamental or direct measurement, put simply, is basic measurement that can be based upon a standard unit preserved in some permanent location (examples are length and weight). Fundamental measurement does not depend on prior measurement of one or more other quantities (eg, compare the measurement of length or weight with the measurement of density). In fundamental measurement, judgments of greater than, equal to, and less than can be made by using a matching procedure and the preserved standard unit. A measurement procedure is not needed unless one wishes to establish an additive scale.

Indirect measurement involves the measurement of one or more other quantities in order to get at the quantity of interest. *Derived measurement* is indirect in the sense that the quantity of interest is obtained from the relationship between two or more other quantities on certain scales under specified conditions. Measures of velocity and density are obtained through derived measurement. Derived measurement is often found expressed in the form of a numerical law (eg, Ohm's law in which $I = E/R$). *Associative measurement* is indirect in the sense that the quantity of interest is assumed or demonstrated to be correlated with the quantity measured. Measures of temperature, by any means, are obtained through associative measurement.

These simplified descriptions of direct and indirect measurement evade the elaborate logic and theory that underlie these kinds of measurement,[11, 14] but the descriptions should provoke some serious thinking about the kinds of measurement we are using in physical therapy. For example: Is the quantity referred to in our literature as "EMG activity" obtained through fundamental, derived, or associative measurement? If the answer is derived or associative, what is the quantity of interest? What *is* EMG activity?

LOGICAL REQUIREMENTS OF MEASUREMENT

Measurement of any kind, whether as described above (fundamental, derived, or associative) or as others may prefer to think of it (range of joint motion, torque, or EMG activity) must meet certain logical requirements. Ultimately, these requirements have to do with establishing scales of measurement.

I have long been an advocate of the following minimal set of logical requirements: Identify a dimension (variable, trait, characteristic, or property) of interest; operationally define the dimension to make it publicly observable; and operationally define two or more categories or units on the dimension in such a way that they are mutually exclusive and exhaustive.[12] These requirements apply equally well to designing and selecting measurement procedures and instruments, and they apply equally well to quantifying and classifying objects, events, and people.

Not everyone agrees that classifying is an act of measurement. One authority states that classifying does not even represent the lowest form of measurement—that it is better spoken of as identification rather than measurement.[15] Another maintains that nominal scales are scales for the measurement of identity and difference only.[14] But a third authority says, "The forming of classes of equivalent objects or events is no trivial matter. An operation for determining equality is obviously the first step in measurement. . . . Without this step, no further measurement is possible."[13 (p26)] The most persuasive reason to consider classifying as measurement comes not from the words of authorities but from the recognition that the nominal scale must meet the same logical requirements as the ratio scale.

There is, of course, the view that measurement has chiefly or only to do with numerals and numbers, and those who hold that view either out of principle or out of preference for a more rigorous mode in which to address the logic of measurement do have some interesting and useful ideas to consider.

Suppose that you have just measured the range of glenohumeral abduction, using the appropriate instrument, and that you have written down the figures 0 to 60°. In assigning those figures to your observation, did you assign numerals or numbers? The question is not trivial.

A numeral is a symbol that may be used to represent a number. Numerals and numbers are said to have a close relationship because they share a common property: possession of a definite order.[11] Campbell starts with this point and builds a set of logical requirements.

Numerals have a definite order by invention, whereas numbers have a definite order that arises from the real properties of things. This real order arises from certain relations between the things that are ordered. These relations are transitive and symmetrical.

The transitive relation is such that if A has it to B, and if B has it to C, then A has it to C. For example, if $A > B$, and if $B > C$, then $A > C$. The symmetrical relation is such that if A has it to B, then B has it to A, assuming that the relation in one direction may be the converse of the other. In other words, if $A > B$, then $B < A$.

For a dimension to be measurable, and for its representation by numerals to convey any information of importance, the objects that differ on that dimension must be related by some transitive symmetrical relation. Recognize that this statement is a necessary condition for at least the ordinal scale. (The nominal scale is removed from consideration as a scale to which numerals may be applied, and rightly so.) Recognize, also, that the statement implies nothing about the magnitude on the dimension. The numerals on an ordinal scale represent order only, not magnitude. The numerals, in this case, have no quantitative meaning. Finally, recognize that any one of the following sets of numerals may be applied to the categories on an ordinal scale: 1, 2, 3, 4, 5; or 1, 3, 5, 7, 9; or 1, 3, 10, 15, 35. If the numerals assigned to a scale are to represent magnitude or quantity, a transitive symmetrical relation is necessary and the so-called laws of addition must apply.

Campbell's First Law of Addition, as it applies to a quantitative scale, is that $1 + 1 > 1$.[11] Consider weight as an example. If we take two bodies each of which weighs "1" and we combine them, the result is "2." For weight, length, time, and other quantitative dimensions, we can demonstrate that $1 + 1 > 1$. But now consider dimensions measured on the ordinal scale: independence in activities of daily living, ratings of reflex response, and strength as measured by the manual muscle test. If we take two bodies or objects each of which is assigned the score of "1" and we combine them (assuming that is possible), the result is not "2." In this case, $1 + 1$ is certainly not something greater than 1, and the First Law of Addition is not satisfied. The application of this First Law of Addition helps to distinguish between a quantity and a quality.

Campbell's Second Law of Addition is offered in place of the commutative and distributive laws: The magnitude produced by the addition of bodies A, B, C, and so forth, depends only on the magnitude of those bodies and not on the order or method of their addition.[11] If this law is not satisfied, we would not find, for example, that $2 + 5 = 4 + 3$.

Campbell also says that there is only one arbitrary element in measurement, that being the choice of the unit. The choice between inches and centimeters is arbitrary; indeed, the choice is determined by convention and not by logical rules. Satisfaction of the laws of addition, as just described, is independent of the choice of the unit.

Another set of logical requirements has been posed by Ellis.[14] A scale of measurement is said to exist only if the following conditions are satisfied: 1) there is a rule for making numerical assignments, 2) the rule is determinative in the sense that the same numerals would always be assigned to the same things under the same conditions, and 3) the rule is nondegenerate in the sense that it allows for the possibility of assigning different numerals to different things or to the same thing under different conditions. Ellis states that the second and third requirements are necessary to ensure that numerical assignments are informative. Note that these same two requirements, considered together, can be satisfied only if the categories or units on the scale are mutually exclusive.

Ellis identified three essential criteria, in addition to the three requirements, for the measurement of quantities: 1) any object that occurs in the order of the quantity represented on the scale must be measurable by the procedure for measuring on that scale, 2) any object that is measurable on the scale must occur in the order of the quantity represented on that scale, and 3) objects measurable on the scale that are arranged in the order of their numerical assignments are thereby arranged in the order of the quantity.[14] Note that the first criterion can be satisfied only if the units on the scale are exhaustive and that the second criterion can be satisfied only if the operational definition of the dimension to be measured is adequate.

The discussion of the logical requirements of measurement has just about come full circle to alternative ways of expressing the requirements first presented, that is, operational definition of the dimension of interest and operational definition of two or more categories or units on the dimension in such a way that they are mutually exclusive and exhaustive.

SCALES OF MEASUREMENT

This discussion of logical requirements made frequent mention of scales of measurement. For very practical reasons, we in physical therapy should acknowledge the four scales of measurement that are widely, though not unanimously, agreed to in other disciplines having a serious literature in measurement, research, and theory. Physical therapy can take legitimate advantage, in both research and practice, of the full range of measurement possibilities provided by the nominal, ordinal, interval, and ratio scales of measurement. There may well be other scales of measurement or other ways of classifying scales of measurement,[14, 15] but these alternatives do not exceed the range of possibilities provided by the four scales just mentioned.

Descriptions of the nominal, ordinal, interval, and ratio scales are readily available in a number of introductory textbooks on research design and statistical analysis. A simplified description, with pertinent physical therapy examples, was published recently.[12] What is not readily available is a discussion of certain important properties and limitations of each of these scales from the standpoint of assigning numerals.

The categories on a *nominal scale* may be designated by numerals, and these numerals may be assigned to people, objects, or events measured or classified on this scale. The numerals so used have no correspondence with numbers in the sense of either order or magnitude, and they cannot be subjected to any algebraic operations. To avoid the possibility of misuse, investigators should probably not use numerals to designate the categories on a nominal scale. Given two objects, A and B, to be measured or classified on a nominal scale, the relation between A and B may be expressed symbolically as either $A = B$ or $A \neq B$; that is, either A and B are identical or they are not, in terms of the dimension represented on the scale.[14]

The categories on an *ordinal scale* frequently are designated by numerals, and these numerals are often assigned to people, objects, or events measured or rated on this scale. The numerals used in this way correspond to ordinal numbers only, and they cannot be subjected to any of the fundamental operations of algebra. Nunnally states the essence of the principle quite clearly: "In the use of descriptive statistics, it makes no sense to add, subtract, divide, or multiply ranks."[15 (p18)] Ordinal scale numerals represent relative order only, not magnitude, yet these numerals are all too often used, in research and practice, as if they had quantitative meaning. Ordinal scale numerals are deceptive. Symbolically, an ordinal "2" looks exactly like the quantity "2." Given two objects, A and B, to be measured or rated on an ordinal scale, the relation between A and B may be expressed symbolically as $A > B$, $A = B$, or $A < B$.[14] To say $A > B$ is not to suggest how much greater A is than B. On an ordinal scale, nothing is known about how far apart the categories are.

The *interval scale* has units. The intervals between the units are assumed to be of equal size, but the units provide no information on how far a measured object is from the real or rational zero on the dimension being measured.[15] The interval scale has no rational or fixed zero, and for this reason ratios cannot be formed of the numerals obtained when using the scale. Nunnally states that interval scale numerals can be subjected to only addition and subtraction; he also states that ratios can be formed of the intervals but not of the numerals.[15] The latter point is a bit tricky. Given several numerals from using an interval scale, $(6 - 2)/(3 - 1)$ would be a legitimate ratio but $6/3$ would not be. Nunnally's view about forming ratios of intervals seems to make sense given the assumption that the intervals are of equal size. Ellis might not agree. Given four objects, A, B, C, and D, to be measured on an interval scale, examples of the permissible relations among the objects would be expressed symbolically as follows: $A > B$, or $A = B$, or $A < B$; also, $|A - B| > |C - D|$, or $|A - B| = |C - D|$, or $|A - B| < |C - D|$.[14] These considerations aside, and notwithstanding the caveat that interval scale numerals can be subjected only to addition and subtraction, researchers in several disciplines, including our own, treat interval scale data as if there were no limitations on the algebraic operations to which those data can be subjected. Fortunately, perhaps, we have few interval scale measurements, temperature measurement and paper-and-pencil tests being among them.

The *ratio scale*, with units, equal size intervals, and a rational or fixed zero, provides numerals that can be subjected to all algebraic operations. Everyone agrees that the ratio scale provides information on absolute magnitude, has a rational or fixed zero, and yields numerals from which ratios may be formed. Ratio scale measurements are, without reservation, measurements of quantity. Given two or more objects to be measured on a ratio scale, the relations among the objects, expressed symbolically, would include the examples mentioned before for the interval scale plus one set of relations unique to the ratio scale: $A = nB$ or $A \neq nB$, where n represents any rational number. Note that these unique relations can be turned around into $A/B = n$ and $A/B \neq n$, where A/B represents a legitimate ratio.

The ratio scale is invariant over all transformations in which the numerals on the scale are multiplied by a constant, that is, transformations of the form $X' = cX$, where c is the constant. An example of this transformation is the conversion of length in inches to length in centimeters: $\text{Length}_{cm} = (2.54)(\text{Length}_{in})$. The invariance of the ratio scale in this transformation consists of no change in the ratio of the numerals, the proportionality of the intervals, and the location of zero.[15, 16] The interval scale, on the other hand, is invariant under any linear transformation of the form $X' = cX + b$, where b and c are constants. An example of this transformation is the conversion of temperature in degrees Celsius to temperature in degrees Fahrenheit: $T°F = (9/5)(T°C) + 32$. The invariance of the interval scale in this transformation consists of no change in the proportionality of the intervals. In the latter transformation, the location of zero changes.[16] In both transformations, the units on the scales change. Recall that the choice of unit is arbitrary, but recognize that the choice is often bound by convention, not by logical necessity. Inches and centimeters are equally valid units of length.

On some occasions, an investigator may decide to transform the data of an experiment. The reasons for making transformations of data and the methods of making them go beyond the scope of this paper, but useful presentations may be found in Keppel[17 (pp556-559)] and in Winer.[18 (pp397-402)] Transformation of data should be made with caution. Some of the methods for transforming data may help satisfy the assumptions underlying selection of an analytic procedure, but at the same time they may transform the scale of measurement in ways that make the scale no longer invariant. The investigator who wishes to transform data may have to choose between two alternative sets of violations.

Satisfying the logical requirements of measurement and developing or selecting an appropriate scale of measurement, as well as applying the scale appropriately, will enhance the reliability of the process. Reliability is a necessary but not a sufficient condition for validity. Reliability and validity are related to precision and accuracy.

PRECISION AND ACCURACY

Precision and accuracy are not synonymous. In the clinical laboratory world, precision is at the center of the concern for quality control. In practical terms, precision has to do with the closeness of the results of repeated measurements performed on the same material under the same conditions. One batch of pooled serum is repeatedly tested by the same device under constant conditions, and the result is expressed as the standard deviation of the repeated measures.[19] The smaller this standard deviation, the more precise the measurement.

Precision depends on sensitivity, that is, the ability to discriminate, and reliability.[9] Accuracy, on the other hand, depends on freedom from systematic error and has to do with the closeness of results obtained in measuring a variable to the true value in the population.[9, 19] Of course, the true value is never known. People in the clinical laboratory world assume that the average value obtained in a number of different settings using the same method is the true value until new information says otherwise.

RELIABILITY AND VALIDITY

The topics of reliability and validity are large, and the literature on these topics is extensive. There are several kinds of reliability and several kinds of validity, and opinions differ on the relative merit of the various kinds of each, on the relative deficiency of various methods for testing certain kinds of reliability and validity, and even on the usefulness of the terms "reliability" and "validity." Simplified descriptions of reliability and validity, with pertinent physical therapy examples, were published recently.[12]

Within the topic of reliability, there is a budding methodological controversy in physical therapy over the use and meaning of the usual product-moment correlation coefficient as an expression of the degree of reliability versus the use and meaning of the analysis of variance intraclass correlation approach advocated by Bartko.[20] I applaud controversy of this kind because the controversy is rooted in theory, including the theory of measurement error,[15 (p172–205)] and perhaps serious discussion of the issue will find its way into our literature.

No measurement can be valid unless it is reliable, but reliability does not assure validity.[15] There are several kinds of validity that can be classified into two major types, judgmental and empirical. *Judgmental validity*, the less useful and less important of the two types for our purposes, includes face validity and construct validity. Face validity is judgment that a test measures what it is purported to measure. The judgment, usually one of expert opinion, is not testable except by expert opinion. Some authors use the term "content validity" instead of "face validity,"[21] but content validity is one of the objectives of test development, usually undertaken with the help of expert opinion, to ensure that the test has face validity after it is developed.[15]

The term "construct" in construct validity refers to an abstract variable that is derived from theory and posed as a hypothesis.[15] A construct is not observable and is not directly testable; its validity is a function of the adequacy of the theory from which it is derived and the adequacy with which the construct represents the observable variables it is intended to explain.

Empirical or *testable validity*, also called practical validity,[21] includes concurrent validity and predictive validity. Both kinds of empirical validity are demonstrable by a correlation of measures obtained on one or more logically antecedent variables with measures obtained on a logically consequent variable (often referred to as the criterion variable). Given a clinically or socially important criterion variable, the search for one or more empirically valid antecedent variables is an attempt to account for the variance of the measures on the criterion variable by the variance(s) of the measures on the antecedent variable(s). The explanatory power of empirical validity lies not in the numerical value of the obtained validity (correlation or multiple correlation) coefficient but in that value squared. An obtained validity coefficient of .60 indicates that 0.36, or 36 percent, of the variance in the criterion variable can be accounted for by the variance(s) in the antecedent variable(s). The example indicates further that 64 percent of the variance in the criterion variable remains to be explained.

In testing human performance, concurrent validity has to do with using the results of performance on a first test or tests to explain the results of performance on the criterion test when the criterion test is given at about the same time as the first test or tests. Predictive validity has to do with using the results of performance on a first test or tests to explain, that is, to predict, the results of performance on the criterion test when the criterion test is given some time after the first test or tests. Studies of concurrent validity are interesting but not nearly as useful or important as studies of predictive validity.

My comments on validity are best summed up by Nunnally's remark: "Strictly speaking one validates not a measuring instrument, but rather some use to which the instrument is put."[15 (p76)]

COMPLEX PROBLEMS

A look at some complex problems in one area of physical therapy measurement can serve as a useful summary of the issues discussed in this paper. The summary will pose problems and questions, not solutions and answers.

The area of measurement is use of EMG for measurement of muscle activity as a dependent variable. I reviewed our literature of the last 10 years and examined the articles that reported this use of EMG for measurement.

Here is my summary observation of the measurement process: Action potentials are emitted, sampled by means of electrodes, conducted or transmitted, amplified, rectified, integrated (in one of three possible ways,[22] not always described), recorded, normalized, summarized, averaged, and then spoken of as "muscle activity," "EMG activity," or "EMG output" expressed in mV/sec or EMG units/sec. What comes out at the end of this process appears to be quantitative.

Here are my questions: 1) What is the dimension of interest here? What is the dimension on which mV/sec or EMG units/sec are the units? 2) Is this measurement fundamental, derived, or associative? 3) Are the logical requirements of measurement satisfied? 4) What is the scale of measurement? Are there any restrictions on the algebraic operations to which the obtained numerals may be subjected? 5) When the data are normalized, what is the scale of measurement? Are there any restrictions on permissible algebraic operations? Is a normalized score of 40 percent twice as much in magnitude as one of 20 percent? 6) What is the reliability of the measurement? 7) To what use is the measurement put? What kind of validity should be demonstrated for this measurement?

The use of EMG for measuring muscle activity appears to be plagued by a sampling problem in the first step of the process. Some investigators normalize their EMG data to circumvent this problem.[23]

My understanding is that the sampling variations diminish the reproducibility of the measures obtained. If the reproducibility of the measures is low, the reliability of the measurement is low. Now, assume that for any dimension of interest (not just muscle activity as measured by EMG) we have a measurement process of known low reliability, that we take two measures (eg, one at "maximal" output and one at something less than "maximal"), and that we then form a ratio of these two measures (the lesser over the greater) and express the result as percentage of maximum. Is the end product, the percentage of maximum, representative of measurement that is more reliable than the measurement used originally?

We have some difficult and complex theoretical, conceptual, and methodological problems in the use of EMG to measure muscle activity. We have some difficult but less complex problems of the same kind in other areas of measurement. Bending our intellectual effort to the solution of those problems will require knowledge of theory and logic and significant developments in both, all of which will contribute to our developing a science of physical therapy.

REFERENCES

1. Campbell SK: Measurement and technical skills—Neglected aspects of research education. Guest Commentary. Phys Ther 61:523-524, 1981
2. White L Jr: Science and the sense of self: The medieval background of a modern confrontation. Daedalus 107(2): 47-59, 1978
3. Boring EG: The beginning and growth of measurement in psychology. In Woolf H (ed): Quantification: A History of the Meaning of Measurement in the Natural and Social Sciences. Indianapolis, IN, The Bobbs-Merrill Co Inc, 1961, pp 108-127
4. Margenau H: Philosophical problems concerning the meaning of measurement in physics. In Churchman CW, Ratoosh P (eds): Measurement: Definitions and Theories. New York, NY, John Wiley & Sons Inc, 1959, pp 163-176
5. Churchman CW: Why measure? In Churchman CW, Ratoosh P (eds): Measurement: Definitions and Theories. New York, NY, John Wiley & Sons Inc, 1959, pp 83-94
6. Shryock RH: The history of quantification in medical science. In Woolf H (ed): Quantification: A History of the Meaning of Measurement in the Natural and Social Sciences. Indianapolis, IN, The Bobbs-Merrill Co Inc, 1961, pp 85-107
7. Gerard RW: Quantification in biology. In Woolf H (ed): Quantification: A History of the Meaning of Measurement in the Natural and Social Sciences. Indianapolis, IN, The Bobbs-Merrill Co Inc, 1961, pp 204-222
8. Kuhn TS: The function of measurement in modern physical science. In Woolf H (ed): Quantification: A History of the Meaning of Measurement in the Natural and Social Sciences. Indianapolis, IN, The Bobbs-Merrill Co Inc, 1961, pp 31-63
9. Kaplan A: The Conduct of Inquiry: Methodology for Behavioral Science. Scranton, PA, Chandler Publishing Co, 1964
10. Caws P: Definition and measurement in physics. In Churchman CW, Ratoosh P (eds): Measurement: Definitions and Theories. New York, NY, John Wiley & Sons Inc, 1959, pp 3-17
11. Campbell NR: Foundations of Science: The Philosophy of Theory and Experiment. New York, NY, Dover Publications Inc, 1957
12. Michels E: Evaluation and research in physical therapy. Phys Ther 62:828-834, 1982
13. Stevens SS: Measurement, psychophysics, and utility. In Churchman CW, Ratoosh P (eds): Measurement: Definitions and Theories. New York, NY, John Wiley & Sons Inc, 1959, pp 18-63
14. Ellis B: Basic Concepts of Measurement. Cambridge, England, Cambridge University Press, 1968
15. Nunnally JC: Psychometric Theory. New York, NY, McGraw-Hill Book Co, 1967
16. Hays WL: Quantification in Psychology. Belmont, CA, Brooks/Cole Publishing Co, 1967
17. Keppel G: Design and Analysis: A Researcher's Handbook. Englewood Cliffs, NJ, Prentice-Hall Inc, 1973
18. Winer BJ: Statistical Principles in Experimental Design, ed 2. New York, NY, McGraw-Hill Book Co, 1971
19. Barnett RN: Clinical Laboratory Statistics. Boston, MA, Little, Brown & Co, 1971
20. Bartko JJ, Carpenter WT Jr: On the methods and theory of reliability. J Nerv Ment Dis 163:307-317, 1976
21. Baggaley AR: Intermediate Correlational Methods. New York, NY, John Wiley & Sons Inc, 1964
22. Tursky B: Integrators as measuring devices of bioelectric output. Clin Pharmacol Ther 5:887-892, 1964
23. Perry J, Easterday CS, Antonelli DJ: Surface versus intramuscular electrodes for electromyography of superficial and deep muscles. Phys Ther 61:7-15, 1981