Diagnosis by the physical therapist has received increased attention in the physical therapy literature. The contributions thus far are in agreement that although physical therapists do not identify disease in the sense of pathology, they certainly can identify clusters of signs, symptoms, symptom-related behavior, and other data from patient history and other testing. These clusters can be labeled as classifications or diagnoses by physical therapists and can guide management of the patient. The purpose of this article is to discuss what has yet to be included in articles about diagnosis: the diagnostic process. We first acknowledge the complexity of the diagnostic process, reviewing the study of clinical diagnosis mostly from the field of medicine, including statistical as well as process-tracing approaches. We next discuss steps we believe are important to consider in order to interface the diagnostic process into entry-level training curricula, urging teachers and mentors of future physical therapists to rethink our emphasis on the problem-oriented medical record’s “SOAP” type of approach as a clinical decision-making format. We next discuss error and clinical judgment and strategies to constructively deal with error in the clinical environment. We urge physical therapists to strive to reach a point at which we can (1) identify and classify patients in such a manner that allows for more efficient treatment management and (2) demonstrate such abilities in peer-reviewed publication form. [Delitto A, Snyder-Mackler L. The diagnostic process: examples in orthopedic physical therapy. Phys Ther. 1995;75:203-211.]

Key Words: Decision making; Diagnosis; Orthopedics, general.

Rose1 encouraged the study of clinical classification in physical therapy, admonishing physical therapists to follow other science-based disciplines that initially focused investigative efforts on “observing and describing phenomena of interest.” He noted that classifying patients would (1) organize a body of knowledge of pathokinesiology, (2) form the basis of clinical diagnosis and movement dysfunction analogous to classification systems of diseases, and (3) establish specific groups for research on the efficacy of treatment and program evaluation.

Since Rose’s publication,1 we have seen numerous other communications related to diagnosis in the physical therapy literature, all of which emphasize the importance to the profession of pursuing diagnosis and classification.2-6 In these communications, a good portion of the discussion centers on the appropriateness of physical therapists diagnosing, and we believe the controversy centers on which of two general definitions are used. The classic medical diagnosis can be defined as identifying a patient’s disease by its signs, symptoms, and laboratory data, and the other general definition, which we believe to be synonymous with clinical classification, entails placing a label on clusters of clinical data. For the purposes of this communication, we will use the latter definition, and we therefore use the terms “diag-
nosis” and “clinical classification” interchangeably.

Sahrmann refined the concept of diagnosis by the physical therapist by noting that “diagnosis is the term that names the primary dysfunction toward which the physical therapist directs treatment.” She not only espoused the further refinement of diagnosis by a physical therapist, but in her communication goes on to further characterize diagnosis as a “prerequisite for treatment.” Sahrmann’s concept of diagnosis refutes the notion that diagnosis is a term for exclusive use by physicians. In a subsequent publication, Rose concurred with Sahrmann by noting that “the objectives of a physical therapy diagnosis are focused on classifying dysfunction rather than disease and are directed primarily to planning and predicting outcome of treatment.”

Jette uses diagnosis and classification synonymously, noting that “diagnostic classification is ... nothing more than a taxonomy.” Jette reviewed diagnosis in the framework of the International Classification of Impairments, Disabilities, and Handicaps (ICIDH) developed by the World Health Organization, and he encourages physical therapists to think of diagnosis in terms of impairment, disability, and handicap diagnoses. Guccione took a very similar approach except that he (1) encouraged using the conceptual framework of Nagi as opposed to the World Health Organization’s ICIDH model and (2) offers a very detailed conceptual framework integrating diagnosis by a physical therapist within the Nagi model. Again, working within the ICIDH and Nagi schemes, Dekker et al. were the first to actually assess the relationship between diagnosis and the treatment physical therapists use, based on a survey of Dutch physical therapists. They concluded that diagnoses based on physical impairments guided treatments that were administered.

All of the authors discussed agree that (1) physical therapists do not identify diseases (in the sense of pathology); (2) diagnosis and clinical classification can be viewed synonymously; (3) movement-related dysfunction can be described by physical therapists through the clustering of data obtained from other sources (charts, other examiners’ reports, laboratory tests), signs, symptoms, and personal characteristics of the patient, eventually leading to a recognizable classification or diagnosis; and (4) the classification can be used to guide the physical therapy management of the patient.

Although few would argue with the virtues of pursuing the area of clinical classification or diagnosis, we note that virtually all the preceding literature has focused on the questions, “Can and should physical therapists diagnose?” and “In what context should the physical therapist diagnose?” The purpose of this article is to delineate what we believe has not been addressed thus far in the discussion and subsequent controversy that sometimes surrounds diagnosis by the physical therapist, that is, the diagnostic process.

The Diagnostic Process

Jette cautions the profession against “bringing the development of classification schemes to closure too quickly in our understandable urge to advance the profession of physical therapy.” Although we may look toward studying diagnosis and classification as advancing the science and body of knowledge of physical therapy, we should not underestimate the complexity of the diagnostic process. In the physical therapy literature, discussions of diagnosis invariably make comparisons with medicine. We would caution that by making such comparisons, we should not presume that medicine or any other discipline has truly mastered the diagnostic process. To quote Eddy:

Whether a physician is defining a disease, making a diagnosis, selecting a procedure, observing outcomes, assigning probabilities, assigning preferences, or putting it all together, he is walking on very slippery terrain. It is difficult for nonphysicians, and for many physicians, to appreciate how complex these tasks are, how poorly we understand them, and how easy it is for honest people to come to different conclusions.

Formal study of the diagnostic process has spurred considerable debate both within the physical therapy profession, including debate from psychologists, economists, decision theorists, statisticians, lawyers, sociologists, and medical specialists (for a review, see Dowie and Elstein). As physical therapists begin to evaluate diagnostic processes pertinent to our profession, there are many available models of evaluation based on recent research and application of the diagnostic process in other fields. We should also keep in mind the limitations imposed by studying a process that is far from exact in almost any of the fields mentioned.

One particularly pertinent approach to studying the diagnostic process is to model the clinician (usually a physician) and the clinical task at hand. Dowie and Elstein describe two distinct approaches that are used to characterize how clinicians make judgments and decisions: the statistical approach and the process-tracing approach. In the statistical approach, the goal is to model the relationship between input (clinical data or cues) and output (the clinician’s judgment or decisions) in the form of a mathematical equation. There is no attempt to model what goes on in the clinician’s head, taking a “black box” type of approach. The process-tracing approach has the reverse aim of attempting to formalize what goes on in the clinician’s head. A spin-off of process-tracing approaches is the knowledge-based or “expert” systems approach.

Whether statistical or process-tracing methods are used, there have been attempts to characterize the diagnostic process, and we will summarize two such attempts. Elstein and Bوردage characterize medical problem solving using strategies that model the clinician and the clinical task into four major categories: (1) cue acquisition, where data are obtained by the clini-
cian by a variety of methods, including history, physical examination, and so forth; (2) hypothesis generation, where alternative problem formulations are retrieved from memory; (3) cue interpretation, where the data are interpreted in the light of alternative hypotheses being considered; and (4) hypothesis evaluation, where the data are weighed and combined to determine whether one of the diagnostic hypotheses can be confirmed and, if not, alternative hypotheses and data collection commence. Another example of a process-tracing approach would be the work of Eddy and Clanton, who examined 50 case reports published in the New England Journal of Medicine and suggested a model with six steps in which clinical problems were solved: (1) aggregation of elementary findings, (2) selection of a “pivot” or pathognomonic finding, (3) generation of a cause list, (4) pruning the cause list, (5) selection of a diagnosis, and (6) validation of the diagnosis.

Elstein and Bordage argue that such reasoning processes shift an ill-defined, open-ended problem (eg, “What is wrong with this patient’s shoulder?”) into a series of better-defined problems (eg, “Is the shoulder pain of muscular or joint origin?”), allowing the clinician to work backward from the diagnostic criteria for each hypothesis to the potential tests and procedures to be conducted in the remainder of the examination. Given the wealth of information that a clinician must process, early aggregation of findings into possible diagnoses is a necessary noise-reducing strategy. To quote Cabot, “To throw open mind’s door and allow all disease to enter into consideration each time that we are called to the bedside is foolish in the attempt and impossible in the performance.” The trick, of course, in formulating initial hypotheses is deciding which initial hypotheses should be considered further versus which hypotheses should be ruled out.

Eddy and Clanton’s work suggests that clinicians begin aggregating elementary findings (eg, any single piece of information about a case) in a hierarchical fashion in an attempt to find a recognizable pattern. In some instances, a key piece of data results in a “pathognomonic” finding; otherwise, clinicians eventually rely on one or two most likely hypotheses, usually formulated early in the data-collection process, and begin a process of confirming and disconfirming hypotheses through additional data collection and testing.

By aggregating initial findings, the clinician searches for a pattern of findings to emerge, makes a judgment about the likelihood of various initial hypotheses, and spends the remainder of the examination gathering data to further confirm these initial judgments. We believe the most important issue to emphasize is aggregation and arrival of early hypotheses, which the majority of time occur very early in the examination process with much of the remainder of the examination process guided by these initial findings.

Although patient evaluation is taught in our professional curricula, we contend that it is performed in a manner that largely ignores the science of diagnostic process. We further contend that procedures related to diagnostic process can be taught and represent some of the most important knowledge we as teachers and mentors can transfer to developing physical therapists.

Diagnostic Process: Interfacing With Present Entry-Level Curricula

In addition to building a solid foundation to the physical therapist’s scientific endeavors into the diagnostic process, physical therapy faculty will need to implement newly gained knowledge within existing entry-level curricula, a task we believe should begin by asking critical questions related to our existing mode of teaching.

The Problem-Oriented Medical Record and “SOAP” Notes

The problem-oriented medical record (POMR) was first introduced as a documentation system with the purpose of organizing a medical record by first listing patient problems in the front of the chart and then “imposing” that clinicians write separate SOAP-type notes related to the problem(s) that were identified. For patients with multifaceted medical problems who were being treated by clinicians from numerous disciplines, the organization imposed by the POMR presumably centered each clinician’s effort toward a coordinated approach (and accountability) to the patient’s problems, thus offering one major advantage in using the system. In reality, however, the POMR system is rarely used in its entirety, perhaps because many patients entering the health care system do not have multifaceted problems and using the entire POMR system becomes burdensome.

One of the offshoots of the POMR system was the SOAP note. “SOAP” is an acronym for Subjective, Objective, Assessment, and Plan, with each heading representing an organizational format for a medical record. Variations of the SOAP format have been embraced by many who teach in our entry-level curricula. In addition to serving as a documentation format, the SOAP model is also used as a clinical decision-making tool, and this is where we wish to focus most of our attention.

We would argue that the diagnostic process is many times stifled when the clinician relies strictly on the SOAP framework for clinical decision-making guidance, and we will illustrate our basis for this statement. The diagnostic process should not be thought of as a simple collation of data from a “laundry list” of examination procedures. The notion that in an ideal clinical situation the clinician should build each patient’s case individually has merit from the standpoint of a classical scientific approach. This attempt to avoid bias probably contributes to the dictum to students to collect all infor-
mation during the history and physical examination before attempting an "assessment." This approach is, however, an unrealistic oversimplification that is not helpful in real clinical situations, where the goal is diagnosis and treatment planning.17 Jensen and colleagues18 described how master orthopedic clinicians were more confident at predicting outcomes than were novice clinicians. The master clinicians attributed this ability to their ongoing interpretation of the information they gather during an examination, an integrative approach that correlates the information they gather with experiential knowledge and consideration of patient needs and a fluid approach to evaluation. Conversely, the novice clinicians "held firmly to their evaluation framework as a basis for decision making."19

The findings of Jemen and colleagues18 confirm our experience with students in most clinical situations, namely that students and new graduates do not operate with the same efficiency as experienced physical therapists. The question that we raise in this article and direct toward teachers and mentors of future physical therapists is, How much of this inefficiency can we attribute to inexperience versus the shortcomings of present entry-level curricula? Although it may be comfortable for us in academia to view any shortcomings of a new graduate as easily addressed through experience, we would argue that future demands placed on physical therapy by health care reform will most likely place a premium on efficiency and that an increasing burden will be placed on entry-level curricula to produce a new graduate who can determine very quickly and accurately the most effective interventions.

Our problem with the SOAP-type note is not related to its use as a documentation format. The accountability provided through systematic evaluation of the relationship between treatment and outcome is enhanced by the SOAP types of documentation, although we believe that a more pertinent approach is available for physical therapists through alternative documentation approaches.19 When the SOAP-type system becomes an evaluation scheme,20 however, we believe there are several shortcomings. First, we believe that the "subjective" component of the SOAP format, whether intended or not, tends to discount all patient-supplied information as lacking adequate measurement characteristics on which to base diagnostic decisions, an assumption that we reject out of hand. More importantly, we believe that a sequential rather than integrative approach to clinical decision making is perpetuated by the emphasis placed on the SOAP format. As a clinical decision-making format, the SOAP process implies that a hypothesis of how a patient will be treated should be attempted only after the history and physical examination are performed. We maintain that generation of treatment hypotheses related to treatment decisions is a process that is continuous with the examination, and often some elements (including an initial hypothesis that may be a diagnosis) may actually precede seeing the patient. This may be especially true if the clinician has a knowledge base of experiential data.

The diagnostic process involves integrating clinical data and experiential and didactic information with decisions being made constantly during as well as after the history and examination. Asking a clinician to check his or her experience at the door and start fresh with each patient disregards the wealth of information and clarity of thought that experience can bring to the diagnostic process. It also disregards human nature. Throughout the history and examination, information is processed and decisions are made concerning what question and what examination procedure to perform next.

Expert clinicians generate alternative hypotheses to test and refine throughout out the history and examination.21 The diagnostic process entails information processing such as hypothesis generation and testing, heuristic searches, and pattern recognition. For more complete descriptions, the reader is referred to an excellent text on clinical diagnosis by Balla.21 We provide brief descriptions and examples.

**Hypothesis generation and testing.** From the beginning of the patient care situation (perhaps as early as looking at a referral slip), a hypothesis can be generated, guiding further data collection in order to test the hypothesis. A patient's chief complaint may be buttocck and posterior thigh pain, resulting in a clinical hypothesis implicating the low back. Further examination, however, may reveal limited passive motion and reproduction of pain at the end-ranges of hip motion, leading the clinician to place less emphasis on the lumbar spine and more on a thorough hip examination.

**The heuristic search.** The heuristic search is defined by Balla as ... a goal-oriented method and to be able to use it one must understand the structure of the problem and know the likely outcomes. Information gathering will then be oriented in such a way that the data can be used to reach the chosen goal.21

For example, the patient with buttocck and posterior thigh pain may upon further questioning reveal that the pain is extreme, occurs at night and is so severe that it keeps the patient awake, and is not affected by movement. All of the symptoms described should raise "red flags" to the clinician, and the "rule of thumb" is that such symptoms may indicate pain of nonmusculoskeletal origin, with referral to medical sources imminent.

**The pattern-recognition method.** The pattern-recognition method, according to Balla, "involves reducing a difficult problem into something that is known or recognized."21 Knowing exactly what is recognized as useful data and what is extraneous is usually attributed to the clinician's experience. The Table illustrates the exhaustive set of data a clinician may collect on the patient with buttocck and posterior thigh pain. An experienced clinician may note a few key pieces of information that represent a pattern. Information is collected, and very early in the examination a pattern is recognized.
In a 65-year-old man, low back and what may be radicular symptoms that are brought on by walking and relieved with sitting may indicate either spinal stenosis or intermittent claudication. The expert clinician quickly decides which questions and tests will be necessary to verify one of the two hypotheses, such as comparing the patient’s response after walking on a treadmill versus riding a bicycle. Pain in the lower extremity that arises from intermittent claudication should be brought on by both the treadmill and the bicycle exercises. Radicular pain of spinal stenosis origin is usually exacerbated with spinal extension postures (eg, as in walking) but is not affected with spinal flexion postures (eg, as in stationary bicycle exercise).22

Just as important as determining necessary tests, the expert clinician must also choose which tests are not important to perform, thus reducing the signal-to-noise ratio in the pattern-recognition process. Clinicians in orthopedic physical therapy exhibit all of these forms of information processing. Consider the process of “clearing a joint” (examining a joint proximate to the area of complaint to rule out involvement of that area.) In examining a patient with shoulder pain, one of the first tests usually performed in the upper-quarter screening examination is full passive motion with overpressure of the cervical spine in all planes. The underlying assumption is that this test may evoke radicular symptoms from the cervical spine. If the radicular symptoms reproduce the patient’s complaints, then further testing of the cervical spine becomes the “rule of thumb.” Further, the clinician has generated a hypothesis that pain in the shoulder can radiate or be referred from the cervical area, and that the cervical motion is a test to rule out the hypothesis. If the cervical motions do not reproduce the shoulder symptoms, then the hypothesis is modified and the examination proceeds with emphasis on shoulder testing. If symptoms are elicited and include shoulder pain similar to the pain that the patient complains of, then a more thorough examination of the cervical spine is indicated. Further testing may include upper-extremity reflex, sensation, and strength assessment to determine whether a “pattern” can be recognized that would implicate cervical nerve root involvement.

Although this scenario may make sense and an upper-quarter screening examination is a classic component of orthopedic assessment and treatment courses, batteries of tests such as the upper-quarter screening examination have not been subjected to any scientific scrutiny. Is the absence of radicular pain resulting from cervical passive motion tests adequate to rule out cervical pathology, or should other testing be performed? How accurately can clinicians judge radicular symptoms? Answers to these questions would begin to address the predictive capability of the test, and would be extremely useful in teaching the upper-quarter screening examination to students. Questions such as these often bring about debate among orthopedic physical therapists but have yet to lead to any peer-reviewed publications.

The teachers of clinical classification or how physical therapists should diagnose are unable to tap a body of scientific knowledge and rely mostly on the authoritative knowledge that guides day-to-day operations in the clinics—knowledge usually obtained through firsthand experience, continuing education, and, in unfortunately few instances, through some classic textbooks.23-27 Instead of teaching and modeling a diagnostic process, we are commonly guilty of teaching to gather data in an exhaustive manner, then come to an assessment and make decisions regarding treatment. We confuse completeness with exhaustiveness and teach a “laundry list” of examination procedures and tests that

---

Table. Exhaustive List of Clinical Data From a Hypothetical Patient With Low Back and Posterior Lower-Extremity Pain

| ● 65-year-old man | ● No symptoms produced with “spring testing” (deep pressure applied over the spinous process at each lumbar level) |
| ● Moderately obese (13.6 kg [30 lb] overweight) | ● Goniometric assessment of lumbar motion reveals 50° of lumbar flexion, 30° of left side-bending, 25° of right side-bending, and 10° of spinal extension |
| ● Pain most severe in posterior calf area bilaterally | ● A slight scoliosis is observable, apex to the left |
| ● Pain to a lesser degree in low back area | ● Isometric testing of flexion and extension is strong and nonpainful |
| ● Absent Achilles tendon reflexes bilaterally | ● While positioned supine, hips at 90° and knees maintained in extension, patient is able to lower both legs, keeping back flat against table until hips are at 0° (lower abdominal strength testing) |
| ● Negative straight leg raise bilaterally | ● Patient can maintain a level pelvis with one-legged stance (negative “Trendelenburg” sign) |
| ● 4/5 muscle power throughout lower extremities | ● Patient has a 0.64-cm (0.25-in) leg length discrepancy (“true” leg length as measured from anterior superior iliac spine to medial malleolus) |
| ● Sitting unlimited | ● Heights of iliac crests are equal (patient standing) |
| ● Cannot stand >10 minutes | ● Heights of posterior superior iliac crests are equal (patient standing) |
| ● Cannot walk >1 city block; walking is limited by calf pain and relieved immediately when patient assumes a sitting posture | ● Passive pelvic translocation (in the frontal plane) has no effect on the patient’s symptoms |
| ● Proprioception testing reveals patient can accurately detect direction and amplitude of passive movements to the joints of both lower extremities | |
we expect the student to proficiently perform and interpret. Consensus opinions on important issues related to both measurement and clinical decision making are not presented, perhaps because we are uncomfortable with exposing our clinical uncertainty and our perception that we, as practitioners, do not practice what we preach.

We contend that students must not only acquire the psychomotor skills to perform examination procedures (in itself a formidable task), but also begin to interpret the results of the examination in light of the patient's functional complaints. As a first step, we strongly believe that fundamental issues of reliability and validity must be moved from the research design courses and become a basic part of clinical courses, especially when specific information related to clinical measures is an ever-increasing part of literature.

Although most students can define reliability, most are unfamiliar with the clinical relevance implicit in the reliability of measurements as well as the manner in which they should ask about reliability. The topic of discussion in orthopedic physical therapy courses should be how a student uses information related to reliability to make clinical judgments. How is reliability judged? When is a measurement so unreliable that it is not clinically useful? What options are available to the clinician should a measurement have a high degree of error (low reliability)? We have found that these types of questions move discussions from an arbitrary, statistically based approach to one of a measurement's degree of error. Statistical error is weighed against clinical meaningfulness. The decision to use a measure is then not based on arbitrary labels of "reliable" or "unreliable" but rather on the ability of a given measure to discern clinical phenomena or change given a known error.

Rather than focus on the shortcomings of clinical measurement techniques by dismissing various approaches based on arbitrary statistical cutoffs, we might consider techniques that avoid clinical disagreement and help students and clinicians learn from mistakes, keeping in mind Sackett and colleagues' words that "the overriding criterion to use when deciding which data to seek is the usefulness of a given piece of diagnostic data to the clinician who seeks it and the patient who generates it." 26 Sackett et al.28 propose several strategies for preventing or minimizing clinical disagreement, many of which we see commonly used by physical therapists at all levels of expertise, including (1) corroborating key findings (eg, repeating key elements of your examination), (2) confirming key clinical findings with appropriate diagnostic tests, (5) reporting evidence as well as inference, (4) using appropriate technical aids, (5) arranging for independent interpretation of observational test data, and (6) applying the social sciences as well as the biological sciences.

Living With Error

To be convinced that living with error is the rule in the clinical environment, the physical therapy student need only look to other clinical disciplines. To once more quote Eddy:

"Whether a physician is defining a disease, making a diagnosis, selecting a procedure, observing outcomes, assessing probabilities, assigning preferences, or putting it all together, he (she) is walking on very slippery terrain."

It becomes difficult for the student to accept that error-free measurement is an unrealizable ideal. Students' comfort with the ambiguity and uncertainty of the clinical world should be facilitated. For example, some aspect of the patient history can be expected to be unreliable.29 Patients often have difficulty communicating the nature of their problems, remembering the specific timing of events, and interpreting the clinician's questions; by the same token, clinicians often have similar difficulties interpreting and remembering patients' accounts of their illness. Yet, we do not ask a student to abandon taking the history of a patient.

Instead, the student learns to focus on confirming and disconfirming data and inconsistencies. The patient with low back pain who claims to have increased symptoms with superficial palpation, axial compression, simulated rotation, and other tests of symptom magnification as described by Waddell 30 demonstrates an inconsistency. On the surface, such information may appear to add error to the clinical situation, yet Waddell offers a mechanism to make such information useful to the clinician. In essence, Waddell has applied the social sciences to the biomedical sciences of medicine, as suggested by Sackett et al.31 Teachers and mentors must continue to identify sources of potential error where they are known and to offer explanations and viable alternatives such as those proposed by Waddell.30

We will next discuss a variety of studies that focus on the reliability of clinical measurements that are directly apropos orthopedic physical assessment. Elven and colleagues32 demonstrated that measurements of rear-foot motion based on the subtalar joint neutral position were so unreliable that clinicians could make absolutely contradictory recommendations for foot orthotics for the same patient. A partner paper33 identified a procedure for minimizing clinical variability as well as uncertainty while applying the subtalar joint neutral technique and interpreting its results. This type of reading and a discussion of its implications are essential to the development of a mature clinician who can deal with measurement error.

The methods proposed by McKenzie27 are among the most widely used techniques for evaluating and treating patients with low back pain. Recent work by Riddle and Rothstein,34 however, has shown questionable reliability when clinicians are asked to place patients with low back pain in one of the treatment categories proposed by McKenzie. Similarly, Kendall et al.35 have proposed that certain relationships exist between postural measurements, yet when put to the test such relationships are not demonstrable.34,35 Should we discontinue teaching the methods proposed by McKenzie and
Kendall? We believe the answer is no. They should be critically analyzed and an attempt should be made to determine the source of error, as well as what steps could be undertaken to decrease the error or to make judgments by taking the error into account. Clinical and academic faculty in orthopedic physical therapy are obligated to contribute to the body of knowledge by modifying or replicating such experiments. We also need to be prepared to abandon such approaches when, through repeated testing in peer-reviewed formats, a converging body of evidence fails to demonstrate the usefulness of such approaches.

Students should also note various commonly used strategies that allow clinicians to deal effectively with error. A solution to an error-prone measure may lie in a simple training session, a better definition, or a plausible explanation to account for a discrepancy in the reliability literature. We mentioned that Elveru and colleagues found substantial measurement error in foot and ankle assessments. Diamond and colleagues, however, found that after a minimal amount of examiner training, the identical foot and ankle measurements showed adequate reliability.

In the case of multiple measures that illustrate one underlying construct, a composite of individual measurements can improve reliability substantially. Potter and Rothstein found that measurements obtained with selected tests of sacroiliac joint function generally had low reliability when assessed separately. Because the results of such tests are commonly used to guide treatment decisions for patients with low back pain, lack of reliability could render such approaches suspect. Because such tests are rarely used individually, however, Cibulka et al. studied the reliability of measurements obtained with a cluster of identical sacroiliac tests and found excellent reliability (κ = .88).

Taken as a whole, the implications of such work are that students must be reminded of their own limitations in interpreting the data that they obtain and that they deserve to know what sort of reliability they can expect from instructors using such clinical measures.

**Decision Making, Information Processing, and Classification**

Although the reliability of measurements obtained for a test is essential, such reliability is not sufficient in the clinical decision-making process leading up to classification or diagnosis. In order to classify, the student must also be cognizant of properties of clinical tests that many times are not a part of the entry-level curriculum. Complete definitions and discussion of sensitivity, specificity, and negative and positive predictive values of tests are beyond the scope of this article, but are covered in some excellent reports. With a few exceptions, we are hindered by the lack of a scientific or even an adequate descriptive body of knowledge related to clinical tests and measures.

Both authors of this article have had the responsibility for coordinating the teaching of orthopedic “evaluation and assessment” courses in physical therapy curricula. The dilemma that we have faced is the absence of an accepted body of knowledge in orthopedic physical therapy assessment that has a rigorous scientific basis. Although there are tests and measures that can quantify impairments such as restricted range of motion or weakness, we lack a systematic set of decision rules that would aid assessment when the relationship between impairment and dysfunction is not straightforward. Fortunately, we are beginning to see an emergence of measurement-based studies, most of which detail reliability of many operational definitions used in clinical situations on patient samples. This is a good beginning, but literature examining other critical information (eg, sensitivity, specificity) necessary for classification is still lacking.

In the absence of adequate primary sources for classification of movement dysfunction, we are tempted to use the textbooks written by what can be best described as the “gurus” of physical therapy. These books present an individual’s approach to evaluation and treatment of patients where validity and effectiveness are accepted out of hand. Should we teach the students a particular guru’s approach and allow for a valid criticism that we are perpetuating dogma. Or should we abandon all major guru-style approaches in the curriculum and leave the student with only basic assessment procedures that characterize physical impairments?

Although the latter approach is more conventional and scientifically sound, we have found that implementing this approach does not serve the students at all when impairments and dysfunction are not directly related, as in the assessment of patients with low back pain. In such cases, a compromise is reached where guru-style approaches are freely discussed and even in some cases formally taught. Instructors are obligated to clearly delineate the limitations of these approaches. The student should be instructed to use functional outcome to measure success or failure of any treatment rather than anecdotal evidence or measurements of impairment. The lack of scientific verification of each approach is clearly articulated to the student. Teaching of limitations can be reinforced by presenting the students possible areas of study that could provide clinical and scientific validation for a particular approach.

**The Status of Clinical Classification in Orthopedic Physical Therapy**

Classification and diagnosis in physical therapy are not encountered for the first time within this article. The discussed writings, however, comprise musings and special communications that fall short of offering tangible methods of assessment and classification to the clinician seeing patients and making decisions about their care.

We believe that theoretical works do provide both a conceptual basis and an adequate justification for the pursuit of clinical classification. Theory, however, must be translated into prac-
tice and reported in a scientifically credible fashion.

The classic textbooks that begin to define a body of knowledge in orthopedic physical therapy can serve as an excellent framework for scientific testing. The studies by Riddle and Rothstein as well as those by DiVeta et al and Lovell et al would not have been possible without the books by McKenzie and Kendall et al. Authorities who document their evaluation and treatment approaches not only provide clear, operational definitions of testing procedures but, more germane to this communication, articulate decision rules necessary for their mode of patient classification. Descriptive work about clinical classification is a welcome addition to the body of knowledge, especially when dealing with data from patients.

As physical therapists, can we classify patients into treatment-oriented entities? The evidence is equivocal. There is some evidence that well-trained clinicians can use examination procedures to reliably classify patients with low back pain and that matching treatments to clinical classifications can result in better management. The amount of training, however, that the examiners have undergone limits the generalizability of these studies to the average clinician. A study of examiners without specific training other than continuing education courses showed poor agreement when clinicians were asked to classify patients into well-defined categories. From a curriculum-development standpoint, such information suggests that formal training in the diagnostic process itself may be beneficial, a stance highly recommended by some teachers in other fields.

Failure to push forward into the scientific, peer-reviewed world with clinical classification and relying only on authoritative knowledge will result in disappointment. We can list criteria for classification based on authoritative knowledge, and we may even be able to get to the point where we can demonstrate that students and clinicians can “classify” patients reliably. If such classifications do not provide meaningful information by guiding the clinician’s management or developing prognostic information, however, we could very well end up with a list of “nonsensical” diagnoses, otherwise referred to as “I say so” diagnoses. Balla describes such diagnoses as labels based on very loose and variable criteria, which in many instances may be highly dependent on the professional group performing the diagnosis. The patient with low back pain and pain in the thigh can be diagnosed by an orthopedist to have degenerative disk disease, by a physiatrist to have low back strain with radiating pain, and by a neurosurgeon to have a bulging disk with referred pain to the thigh. As physical therapists, we should keep in mind that should we attempt to classify in a similar arbitrary manner, we will only serve to add to the list of nonmeaningful diagnoses. Alternatively, we should strive to reach the point at which we can (1) identify and classify patients in such a manner that allows for more efficient treatment management or improved prognostic ability and (2) demonstrate such abilities in peer-reviewed publication form.

References

31 Elven RA, Rothstein JM, Lamb RL. Gonio-
metric reliability in a clinical setting: subtalar
and ankle joint measurements. Phys Ther.
32 Elven RA, Rothstein JM, Lamb RL, Riddle
DL. Methods for taking subtalar joint
33 Riddle DL, Rothstein JM. Intertester reli-
ability of McKenzie's classifications of the type
of syndrome present in patients with low-
34 DeVita J, Walker ML, Skibinski B. Relation-
ship between performance of selected scap-
ular muscles and scapular abduction in stand-
35 Lovell FW, Rothstein JM, Pesontos WJ.
Reliability of clinical measurements of lumbar
lordosis taken with a flexible rule. Phys Ther.
36 Diamond JE, Mueller MJ, Delitto A, Sinac-
core DR. Reliability of a diabetic foot evalua-
37 Potter NA, Rothstein JM. Intertester reli-
ability for selected clinical tests of the sacroiliac
38 Gibulka MT, Delitto A, Koldehoff RM.
Changes in innominate tilt after manipulation
of the sacroiliac joint in patients with low-
39 Klein AB, Snyder-Mackler L, Roy SH, De-
luca CJ. Comparison of spinal mobility and
isometric trunk extensor forces with electro-
myographic spectral analysis in identifying
40 Cooperman JM, Riddle DL, Rothstein JM.
Reliability and validity of judgments of the
integrity of the anterior cruciate ligament of
the knee using the Lachman's test. Phys Ther.
41 Delitto A, Gibulka MT, Ehrard RF, et al.
Evidence for use of an extension-mobilization
category in acute low back pain syndrome: a
prescriptive validation pilot study. Phys Ther.
42 Delitto A, Shulman AD, Rose SJ, et al.
Reliability of a clinical examination to classify
patients with low back syndrome. Physical Ther-

---

**Call for Reviewers**

*Physical Therapy* is currently seeking qualified individuals to serve as manuscript reviewers. Reviewers should have:

- Extensive experience in area(s) of content expertise
- Experience as authors of articles published in peer-reviewed journals

Familiarity with peer review is essential.

If you are interested in becoming a reviewer for the Journal, please send a cover letter and a copy of your curriculum vitae to:

Editor
*Physical Therapy*
1111 North Fairfax Street
Alexandria, VA 22314-1488

Interested in becoming involved, but not sure you have the
time to review manuscripts? The Journal is also looking for article
abstracters and book/software/videotape reviewers. Send us a
letter expressing your interest and stating your general areas
of expertise, along with a copy of your curriculum vitae.
We look forward to hearing from you.