

# Studies of Reliability: Show Me the Question

Richard P. Di Fabio, PhD, PT  
Editor-in-Chief

---

**T**he obvious question is *not* the most important question when it comes to the study of reliability. The inspiration to become a clinical scientist must extend beyond a concern about the reproducibility of measurements. In view of the proliferation of reliability studies in our profession (a phenomenon that seems to be directly related to the growing number of physical therapy academic programs requiring research projects from their students) we need to re-evaluate our focus on the study of measurement reproducibility.

Confirming reliability should be a precursor to investigating a larger issue. Many studies of reliability do not qualify as self-standing research reports because the most important question is usually missing. Why should we care about the reproducibility of KT-1000 arthrometer measurements if clinical questions involving knee function cannot be answered using this device? Will anyone care about "reliable" force plate measurements if we do not also show that the signals obtained from the plates will provide information about the mechanism of pathological gait or poor athletic performance? Why are we interested in studying the reliability of the Oswestry disability questionnaire or the Roland disability assessment if the scores on these tools do not help us answer questions about the effectiveness of treatment for patients with low back pain?

The relevance and importance of reliability studies as a form of primary research depends not only on the questions asked, but on the population studied. The investigations led by Watson and Powers that appear in this issue are important because these authors answer meaningful clinical questions by testing subjects who had patellofemoral dysfunction. Their recommendations directly pertain to patients with this disorder. I have seen too many studies, however, that evaluate the reliability of clinical tests by studying nonimpaired individuals. This does not help us understand the complexities of measurement in the clinic because the results of reliability studies apply only to the population tested. Pain and altered joint structure can influence the testing position, the effort produced during the test, and the response to repeated test applications. If patients who have specific musculoskeletal impairments are not included in a study of measurement reproducibility, then the reliability of the tool does not apply to these people.

Branding a test "unreliable" will usually spell the end to that assessment procedure. But what happens when treatments that were based on unreliable assessments appear to benefit patients? Commentaries in this issue by Dr Scott F. Dye and Ms Jenny McConnell point out that the relief of anterior knee pain that results from using the McConnell taping techniques should justify the use of taping procedures in the clinic despite the findings of poor reliability of patellar position measurements. Dye and McConnell contend that studies evaluating the reproducibility of patellar position fail to acknowledge that taping can improve a patient's well being. Perhaps measures of patellar orientation need not be precise to identify a taping procedure that will reduce the patient's symptoms. Should we abandon the use of a

potentially effective clinical procedure simply because the determination of patellar alignment prior to the treatment was not reproducible? On the other hand, how can we determine which taping technique to use if we cannot reliably estimate the position of the patella?

The commentaries by Dye and McConnell and the responses by Watson, Powers, and their colleagues present us with a paradox between the assessment of reliability and the selection of effective treatments for patients with patellofemoral pain. Do we accept the finding that patella position assessment is unreliable but continue to use taping as a treatment for patellofemoral pain syndrome?

The reproducibility of any clinical assessment is important, and researchers must establish the reliability of their measures. The study of reliability, however, must be surrounded by a description of why the measure is important and what influence the measurement has on determining the nature and scope of treatment. If the information gained from an evaluation tool has no meaningful bearing on treatment, then the assessment of measurement consistency becomes an irrelevant exercise in statistical method. The reliability of clinical assessments cannot be studied without people who have impairments, nor can it be discussed without answering questions of consequence.