COMMENTARY / COMMENTAIRE

Clinical and Scientific Considerations in Progress Monitoring: When Is a Measure Too Long?

Barry L. Duncan
The Heart and Soul of Change Project, Jensen Beach, Florida

Robert J. Reese
University of Kentucky

In their thoughtful commentary in reaction to the special issue of Canadian Psychology on progress tracking (2012, Vol. 53, No. 2), and specifically to the article by Duncan (2012), Halstead, Youn, and Armijo (2013) posed a useful psychometric question regarding how brief is too brief when considering progress measures. They suggest that measures should be of sufficient length to provide reliability and validity but provide no definition of what constitutes sufficient reliability or validity. Moreover, Halstead et al. overlook the important clinical issue of feasibility, whether the measure will be routinely used by front line clinicians. We assert that there is no doubt that the increased reliability and validity of longer measures likely result in better detection, prediction, and ultimate measurement of outcome, but suggest that empirical investigation is required to determine if these differences are clinically meaningful and offset the low compliance rates. We also assert that while empirical investigation is required to determine how brief is too brief, the answer to the question regarding when a measure is too long is simple: When clinicians won’t use it.

Keywords: PCOMS, reliability, validity, feasibility, ORS

In their reaction to Duncan (2012), Halstead, Youn, and Armijo (2013) pose a useful psychometric question regarding how brief is too brief when considering progress measures. They suggest that measures should be of sufficient length to provide reliability and validity but provide no definition of what constitutes sufficient reliability. They asserted that while empirical investigation is required to determine if these differences are clinically meaningful and offset the low compliance rates, the answer to the question regarding when a measure is too long is simple: When clinicians won’t use it.

Keywords: PCOMS, reliability, validity, feasibility, ORS
mentary, in many ways, provides a ready explanation for the divide. They suggest:

...we think that where it is possible to use more reliable measures that
give us normative information, we should attempt to use them...The
more reliable measures are better at detecting early improvement and
more importantly early deterioration and allow us to track change in
a scientifically meaningful manner. (p. 84)

Although “more reliable” is not defined, “where it is possible” represents the crux of the issue, and it is where the division between research and practice occurs. Or in other words, can the science of measurement be feasible for everyday clinical use? The brevity of the ORS, with its attending lower reliability and validity (although we contend are far from unacceptable), makes a difference, because, as is news to no one on the front lines, and especially in the public sector, the number of forms and other oversight procedures has exploded. Few have the time to devote to the repeated administration, scoring, and interpretation of lengthy measures—feasibility is critical.

Low compliance rates are the most frequent result of longer measures. For example, comparison of two similar sites, one implementing the ORS and the other the OQ revealed a compliance rate for the ORS of 86% at the end of one year, and despite ongoing encouragement, the use of the OQ was just 25% (Miller, Duncan, Brown, Sparks, & Claud, 2003). Furthermore, longer measures often wind up being used as periodic or prepost measures, which result in poor data integrity, not representative of actual practice. For example, a benchmarking study conducted in a managed care setting requiring the 30-item OQ at the first, third, fifth, and every fifth session thereafter lost over 55% of the data for lack of two data points (Minami et al., 2008). Similarly, a study of the CORE 34 resulted in only 9,703 clients with pre- and post information from a database of over 33,000 (Stiles, Barkham, Connell, & Mellor-Clark, 2008).

Measures that are perceived as too long by psychotherapists prevent many from even considering monitoring outcome. For example, in reaction to a managed care company’s introduction of a 30-item OQ, the New England Psychologist (Hanlon, 2005) reported that providers complained about its length and frequent administration. The response by clinicians was so severe that the State Psychological Association president said, “I have never seen such negative reaction from providers” (Hanlon, 2005, p. 11).

Intimately related to feasibility is the issue of the utility of the feedback—whether the measure has an intended clinical use to improve the effectiveness of rendered services. Most outcome measures are used primarily as prepost and/or periodic assessment devices. Such instruments measure program effectiveness but are not feasible to administer frequently, and therefore, they do not provide real-time feedback for immediate treatment modification before clients drop out or suffer a negative outcome—in short, they are not clinical tools as much as they are management or oversight tools. The ORS was designed as a clinical and outcome tool to provide real-time feedback to improve the effectiveness of services and as a way to measure outcomes.

Perhaps this speaks to the normative versus communicative distinction made by Halstead et al. The communicative aspects of the ORS are critical to enhancing outcomes, but the normative aspects provide the credibility for the discussion. There are now over 400,000 administrations of the ORS resulting in algorithms for expected treatment response. It is not surprising that the trajectories are not unlike those reported by other outcome measures.

There is an unfortunate lack of data from where most mental health services are provided—in public behavioural health settings. In many ways, the lack of available data in the real world speaks to the very heart of the divide between research and practice. Wolfe (2012), in a clever dialogue between his researcher and practitioner selves, suggested that practical outcome tools for everyday clinical practice, like the ORS, can serve to build the bridge between research and practice.

There is no doubt that 45 items, 30 items, or even 19 items is better than 4 items, and that the increased reliability and validity of longer measures likely result in better detection, prediction, and ultimate measurement of outcome. But how much better is the reliability and validity and more important, how much better is the detection, prediction, and ultimate measurement of outcome? Are these differences clinically meaningful, and do they offset the low compliance rates and resulting data integrity issues? These are the questions that require empirical investigation to determine how brief is too brief.

But when is a measure too long? The answer is simple: When clinicians won’t use it.

Résumé
Dans leur réaction à Duncan (2012), Halstead, Youn et Armijo (2013) posent la question d’ordre psychométrique suivante : qu’est-ce qui constitue une mesure trop courte pour l’évaluation du progrès? Ils suggèrent que les mesures doivent être suffisamment longues pour assurer fiabilité et validité, sans toutefois offrir de définition de ce qui constitue une fiabilité ou una validité suffisante. En outre, Halstead et al. passent sous silence l’importante question dans le domaine clinique qu’est la faiabilité, à savoir si la mesure sera couramment utilisée par les cliniciens de première ligne. Nous affirmons qu’il ne fait aucun doute que la fiabilité et la validité accrues de mesures plus longues donneraient probablement lieu à de meilleurs résultats en ce qui a trait à la détection, à la prédiction et à l’évaluation finale des résultats, mais nous suggérons qu’il faut avoir recours à l’enquête empirique pour déterminer si ces différences sont significatives sur le plan clinique et si elles compensent les faiblesses tels de conformité. Nous affirmons de plus qu’il faut recourir à l’enquête empirique pour déterminer ce qui constitue une mesure trop brève, la réponse à la question « Quand une mesure est-elle trop longue? » est simple : lorsque les cliniciens ne veulent plus l’utiliser.

Mots-clés : PCOMS, fiabilité, validité, faiabilité, ORS.

References


Received February 14, 2013
Accepted February 21, 2013