INTERNATIONAL
CENTER FOR
CLINICAL
EXCELLENCE

Responding to New Scientific Objections to the ORS

By Dan Buccino
With Jason Seidel and Scott Miller
ICCE
1 October 2012

After a brief trial using the much longer and more cumbersome OQ-45 (Outcome Questionnaire 45), I have been using the ORS since 2003, and requiring that the students in training in AOP use the ORS with all their patients as well. As such, I am familiar with many of the objections to the ORS which appear to involve three main themes: "A measure that *looks* so brief and simple can't possibly be meaningful…", "How can patients really give honest feedback to their therapists sitting right in front of them?", and, "Isn't it really just a patient satisfaction measure and what do *they* tell you anyway?"

Often these first-line objections obscure a more fundamental discomfort with formal patient feedback and transparency in treatment which may impact on professionals' sense of autonomy, however, these first-line objections can generally be addressed succinctly.

"Sure, the measures are brief. That's the point. How useful can a measure really be if it is too cumbersome to be used regularly and which does not offer real-time feedback?"

"Sure, it's tough for patients to give honest feedback to their therapists about how unhelpful they've been. That's why we must be so sensitive to even the slightest slippage below the cutoff on the SRS, and to ORS scores which are not moving."

"Well, no, it's not a patient satisfaction measure. The ORS is an outcome measure of subjective well-being and the SRS is a treatment alliance measure."

Patients can be very "satisfied" with their therapists and yet not do anything differently as a result of their treatment, but in order to have a good outcome in treatment there must be a good working alliance. As Scott Miller has often said, "Outcome must hold process on a leash."

Over the past decade, for variations on the objections above, I have met resistance in gaining wide-spread acceptance for use of the measures. Recently, however, a new set of objections arose which were at once baffling and yet exciting, because they were so unusual and rigorous.

The first new question was whether the ORS actually was an objective, valid, and reliable outcome measure by itself, or if it was more like a "vital component" of a larger intervention (Feedback-Informed Treatment, FIT). The vital component question was interesting because I have been an avid follower of the work of Peter Pronovost, who pioneered the use of simple checklists to reduce ICU infections, to improve surgical procedures, and in other settings. His checklists were not literally engaged in washing people's hands or actually doing the surgery, yet they were a "vital component" in ensuring that everyone was, say, washing their hands, in improving outcomes, and reducing morbidity, mortality, and costs.

In submitting a PCORI grant application recently, I was asked if the ORS was a measure of an intervention or an intervention itself. If we get funded, we will be able to answer that question very rigorously because we designed our study to demonstrate that it is both. But, given that, initially, I was stumped about that question (Is the ORS an outcome measure or an intervention?), I consulted Jason Seidel, PhD, a Senior Associate of the ICCE, and Scott Miller, PhD, a founder of the ICCE, who know well the psychometrics of the ORS.

Dr. Seidel's lightly-edited responses follow:

"The ORS IS an outcome measure by itself. It measures subjective well-being. AND, the ORS is part of FIT (though not required in FIT; any outcome measure could be used as part of FIT).

"The systematic tracking of outcomes and collecting both outcome feedback and process feedback (i.e., FIT) is certainly an EBP (it practically defines EBP! And has been designated as an EBP by SAMSHA.) regardless of the system of feedback one uses (OQ, CORE, SRS, WAI, ORS, HAQ, "ALERT", "SIGNAL", "Clinical Support Tools", etc).

"The ORS (i.e., outcome assessment) is both a measure providing data AND it's an intervention affecting the data themselves (Why do we have to choose one or the other?). Lambert and colleagues amply demonstrated this years ago when they looked at the differences between patients simply providing the data--filling out an outcome measure (but not giving it to their therapist), and clinicians actually seeing the results so they could use it as feedback, providing the data also to patients, and introducing 'signal' advice alerting clinicians to potential risks to the clinical outcome. Lambert et al found that the more data and prompts to action therapists received, the better their patient outcomes, and that patient outcomes improved further when patients themselves were aware of their own feedback responses and trajectories of change in treatment."

After further discussion, my collaborators were willing to accept that perhaps the ORS could be a vital component of a FIT intervention, but not that the ORS could be an objective measure of outcome on its own.

Again, I consulted Jason Seidel, whose response follows:

"So, we can accept it as an intervention, because we can't understand how it can simultaneously be part of the intervention and the measurement of the outcome of the intervention...? We're okay with a FIT intervention even if there's no measure of that intervention's worth? It sounds like the vaunting of EST/EBP as a theoretically universal good over and above the actual impact of the EST on the real people served in an actual clinic. It's scientism over science.

"This reasoning seems a bit bizarre to me: on the one hand, there is doubt about the content/construct validity as an **outcome** variable so it has less value, but on the other hand your judgment that it improves a **process** variable gives it value? Why is it good that you **judge** that it improves the alliance if their judgment is that there's no valid measure of its **actual impact** on the outcome of service?"

The second, new question had to do with the psychometrics of the ORS itself. My collaborators had actually read the supporting literature and were asking legitimately rigorous questions.

For instance, since the measures were developed by Scott Miller and Barry Duncan, and since many of the studies on the ORS were done by them, doesn't that introduce some bias into the evidence? (The psychotherapy outcome literature is full of this sort of bias. FIT can now be dismissed as being biased the way many of the EBPs have been shown to be biased since many of the RCTs showing benefits of a particular model were actually done by the developers of particular models of treatment under study.)

A further question ran as follows. Since we have been using the measures with a generally hard-to-reach, urban population of patients with enduring mental illness, and had proposed to continue to do so, they asked whether the measures had been adequately content- and construct-validated with a similar population. Although the ORS was derived from the OQ45, and the OQ45 had been validated on a clinical population, in the initial ORS and SRS validation studies there was concern that the ORS had only been given to college students and not to a clinical population, and that ORS scores in a clinical population were not compared to "gold standard" OQ45 scores in a clinical population.

My collaborators were willing to grant that there is substantial face validity to the ORS as a global measure of distress and subjective well-being, but were concerned about its content and construct validity in a clinical population, especially a clinical population with enduring mental illness, as opposed to an EAP population or a family services agency population or a population of couples in treatment or a European population of holocaust survivors, which were some of the clinical populations where the ORS had been validated.

"The Campbell & Helmsley (2009) study (see especially table III) does demonstrate concurrent validation of the ORS on a clinical population of patients in a clinical setting referred for mental health care from a health care setting. Anker et al (2009) also demonstrate construct validity, " reported Jason Seidel of the Icce.

Breukers et al (2010) in a largely untranslated Dutch study did use the ORS in a clinical population with patients with mental illness and found that the ORS correlated very well with the SCL-90 with a reliability alpha of r= -.63 (with p<.05) on SCL total score with ORS total score. This study which demonstrated construct and concurrent validity in a clinical l population of mental health consumers led to the ORS being designated a prime outcome measure for psychological treatment in the Netherlands.

Again Jason Seidel continues:

"There is perhaps too much emphasis in the literature on Cronbach's alpha as a measure of reliability. We are often falling short in the research to nail down these important issues in a smart way. I like the criticism—we need more rigor and accountability. I agree that we need to answer questions about validity and reliability and many of these questions have NOT been properly answered. But science is a progression of incremental knowledge.

"The latent variable in the ORS is (subjective) well-being. There is actually no way to measure that or any other subjective experience. There are always intervening variables and huge amounts of error, misdirection, bias, and bad assumptions. Should we just give up, then?

"Remember Plato's *Republic*? We only get at shadows of Well-Being. We use scientific methods to increase the closeness of the shadows to the Truth. That's all we can do. We've scratched the surface of concurrent and criterion-related validity. There's more to be done. I think the "outcome measure" versus "intervention" forced dichotomy is a limitation of our conceptual ability. It's BOTH.

"Part of our dilemma here may reflect a diminishment of the nature of the instrumental measurement of subjective experience, no sensitivity to the reality of time cost, or how one can feasibly estimate the clinical change of a wide, shifting, overlapping, subjective range of clinical phenomena in the brief amount of time that actually makes the whole venture possible.

"This is what surprised me about (the very prominent EBP proponent) David Barlow's comments on a recent American Psychological Association panel - that he seems to now have a strong grasp that the local impact of the EST/EBP must be measured, and it should not be a faith-based process (faith in a local result based simply on the existence of a distant result when we KNOW that the larger variable--percentage-wise--is the here-and-now therapist, not the therapeutic method). Barlow gave as one example Atul Gawande's story about the radically different results with Cystic Fibrosis among two different hospitals that monitored their outcomes; one simply did a better job implementing the standard care (EBP) that everyone at all hospitals was trained in.

"Here is Barlow's quote from his discussion at the APA panel:

> "*One of the issues--and I find this particularly true for those of us in psychology--is* ***we all sit around and wait for the perfect measure****. We say 'Well, I don't know; that one's not quite right,' and 'I know psychometrics and, well, the reliability could be a little higher or this could be or that could be…' or 'After all it's just self-report and we know the problems with self-report' and that kind of comment.* ***But we just can't do that. We have to go with something****, and in the doing--in the using--is when it's going to get better by our shared experience.* ***We can't sit back and wait for the perfect measure,*** *I don't think. In terms of these two types of research, we need both kinds of research.* ***We need the clinical utility or practice research****; and we need the efficacy studies. The efficacy studies can establish the causal relationships.* ***The practice-based research can tell us: Does it make any difference whether you do this or not? Is it going to be useful? Is it going to be feasible?*** *Is it going to be generalizable?"*

This is precisely what the work of Peter Pronovost seems to point us toward. That we need to do more healthcare delivery and services research, more comparative effectiveness research, more outcomes optimizing research, and more applied research on the discovery of practical know-how.

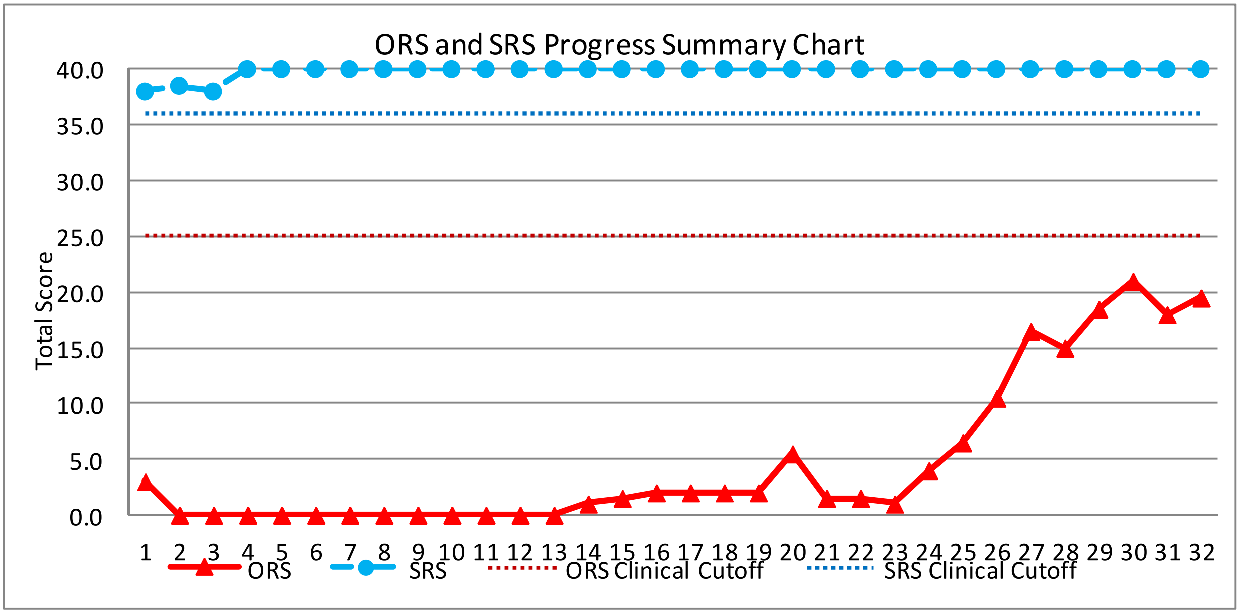
Utilizing highly feasible measures like the ORS with robust face-validity and adequate content- and concurrent-validity, can help us generate more practice-based evidence of our own effectiveness, rather than resisting the call to transparency or arguing about the psychometrics of the measures.

A century ago, surgeons didn't think it was necessary to wear gloves in the operating room. After William Halsted at Hopkins was able to demonstrate that simply wearing gloves dramatically reduced post-operative infections, morbidity, mortality, and cost, OR staff began wearing gloves. Wearing gloves doesn't help you perfect your technique on a pancreaticoduodenectomy, but it does help more patients survive them. The ORS is like a surgical checklist or surgical gloves: they don't do the therapy, but they are a simple way to gauge effectiveness and identify patients at risk of null outcome (or post-op complications).

Given Liz Pfoh's summary recommendations for time and resource efficient outcome measures, which are clinically feasible and meaningful, coupled with the need to facilitate low-cost data entry and analysis, the ORS meets the standard of reliability and validity, and utility and feasibility, and MyOutcomes.com offers an elegant solution to minimize paper data collection and entry. MyOutcomes has extended its final offer of a free trial for us only through October 2012. Now may be the best time to build on our AO results pilot project and utilize the its more robust features of MyOutcomes, including especially its capacity to flag cases at high risk for null outcome.

I have also attached a staff member's reflections on using the ORS for a more qualitative commentary on its utility.

The Use of the ORS and SRS from a Therapist's Perspective



This is the progress summary chart for "Ms. H" over 32 sessions. Sessions span from 1/4/12 to 8/16/12.

Case of "Ms. H"

Ms. H is a 29 year-old, single African American woman who has been a patient at JHBMC CPP since June of 2010. Both of the patient's parents were IV drug users and each died of AIDS, her father when she was in 5th grade, and her mother when the patient was 19 years old. Ms. H completed some college, but dropped out around the time of her mother's death. Pt. is not working and has been on disability since 2008 and lives with her grandmother and younger sister. Ms. H's psychiatric treatment began in 2000 following a suicide attempt. Pt. has had a number of psychiatric admissions for suicidal thoughts or attempts in the years following and has had a number of outpatient treatment providers. Pt. is currently coming to therapy on a weekly basis and attends Bayview's PRP. She has been diagnosed with the following: MDD, recurrent, severe with psychotic features, Eating Disorder NOS, and Borderline Personality Disorder. Ms. H was transferred to this therapist in December 2011 following a 4-month stay in IOPA after previous therapist left following the birth of her child.

The ORS

When I started working with Ms. H in December she would consistently rate herself low on the ORS. At the time she would report "low mood" and express a lack of motivation. In March 2012, Pt. noticed an increase in suicidal thoughts and requested to go back to intensive outpatient services, but instead agreed to meet with this therapist twice a week. By the end of March, Ms. H was beginning to report increased stability and had been making more healthy decisions. This therapist had pointed out that Pt. continued to report "0" across the 4 subscales and asked whether she could report some improvement. Pt. increased her rating to "1" on the Interpersonal Scale and Therapist and Pt. discussed why that was. By mid-April, Ms. H had begun a new relationship, had started going to the gym, and had expressed a desire to return to community college. This therapist pointed out that she had clearly made improvements and talked to Pt. about how she rated herself. Ms. H's ability to notice and rate her own progress had become an important topic in therapy as it is intertwined with her self-concept and negative thinking. It also became relevant to the discussion about what being "well" means. Ms. H returned to weekly meetings, has increased her participation at PRP, and is in the process of preparing for a work program. Most notably, she expressed that she has found her will to "live" and has become more motivated to pursue a life where she is "independent."

Feedback-Informed Treatment (FIT)

As a new therapist with an interest and openness to research, my supervisor asked that I participate in a pilot study of FIT using the Outcome Rating Scale (ORS) and Session Rating Scale (SRS) during my sessions. At first, I was apprehensive and skeptical about administering the SRS because I was concerned that it might make patients uncomfortable to rate my therapy skills, but I soon created a different thought process around it. For one, it was OK for patients to rate the sessions because it let them know that I wanted their feedback and that I could handle it. Also, I explained that we both had a responsibility to creating a helpful session. While my job was to offer a non-judgmental and safe environment, it was the patient's job to bring up important and relevant topics. Therefore, the SRS only added to the discussion about how we could both improve the quality of the session. Plus, when a patient rated the session less favorably I felt like I was able to model how to handle criticism, and I think this helps to reinforce the trust between therapist and patient. Overall, patient feedback has been positive or neutral. Many patients take it very seriously. I have one patient in particular who uses it as a time to reflect on how she made use of the session.

While the SRS provides reliable and useful feedback, I have found the ORS to be even more valuable to the therapy process. From the very first session, the ORS ratings provide a baseline measurement of patients' perceived functioning and it leads the way to a variety of different conversations pertaining to patients' goals. When I introduce it, I explain that the instrument is intended to help us measure progress and understand what is helpful. Sometimes patients have difficulty evaluating their progress, especially when they experience chronic dysthymia, and the ORS immediately begins to help patients learn to recognize improvements. On the other hand, some people consistently rate themselves high, which provides information about their readiness for treatment, comfort with discussing negative topics, and perhaps comfort in the therapeutic relationship.

I also find that the use of the instrument is different for each patient. For some, we barely talk about it in the session, but it may provide a consistent "beginning" and "end" to the session. For others, there have been times when it has been helpful to provide patients with the graph demonstrating their progression in therapy. I am also not overly strict with my administration of the instrument. If someone comes into therapy in crisis and forgets to complete it I may not even mention it to show that I am flexible to their needs at the time.

In addition to its usefulness in the therapeutic process, viewing the data has been helpful to my overall growth as a clinician. If someone consistently reports low scores on either measure, I am likely to bring that patient up in supervision. The SRS has also been shown to predict drop out, which is very useful as a clinician. Additionally, I believe that being able to demonstrate improvement in my patients provides more information about my skills as a therapist than other current measures such as the OMS or VOS, which fail to directly measure therapeutic outcomes. While it may take an extra few minutes each session, I believe that these measures provide a valuable service to patients who expect the highest quality treatment. Johns Hopkins, after all, prides itself for its excellence in education, research, and practice, and I believe the use of measurements such as the ORS and SRS satisfy this mission the way that current practices do not.