Reviewer's report

Title: Instrumentation Issues in Implementation Science

Version: 2 Date: 27 March 2014

Reviewer: Gary Bond

Reviewer's report:

Overall assessment of the revised paper. I now understand that this is a paper in one of the standard manuscript categories for the journal and that the authors and the editor had preliminary discussions over email prior to submission. I am unsure what the criteria are for a debate paper – as opposed to a scholarly review – but I am assuming that the main point is to express a point of view (like an editorial), without being held to the standards of scholarship required of a review article.

Overall, I approach the issues of implementation from a different standpoint than the authors, who appear grounded in traditional academic psychology. In my experience, attitude scales, which the authors showcase, have had limited utility for guiding implementation of evidence-based practices in clinical practice. As noted in my last review, I believe a balanced presentation of measurement options would be more helpful, but I also recognize the authors are presenting their own point of view (informed by their survey) and are not necessarily trying to represent the opinions of the entire field.

Survey. The authors have improved their paper by providing more details on the survey used as one of the foundations for this paper. That said, the survey used a series of ad hoc questions and was administered to a sample with unknown biases and modest return rate. The survey instrument itself was a brief list of items that were rated on a Likert scale. I would characterize the findings from the survey as rather superficial.

Compulsory Revision: Please include a section on study limitations of the survey.

Issue 1: Looks fine.

Minor Essential Revision: Please explain what a Wiki is.

Issue 2. The authors note that “the focus on reliability assessment is most often on internal consistency.” This is probably true because of the implementation field’s overreliance on self-report and the ease of computing internal consistency for multi-item checklists. The issue should not be on what is easiest to do; the issue should be on what is needed. For observational measures, interrater reliability and test-retest reliability are the primary methods for establishing reliability, as the authors note.

Discretionary Revision: Instead of emphasizing internal consistency, my recommendation would be to emphasize assessing reliability using the method most appropriate to the mode of measurement.
Compulsory Revision: The authors state that the EBPAS is “an exemplary instrument (with respect to its psychometric properties).” Please state which psychometric properties are included. In particular, has the EBPAS been found to predict (non-self-report) implementation or clinical outcomes, which is probably what many respondents in their survey mean when they say they are looking for practical instruments?

Issue 3. The diagram is quite abstract. Most of the suggestions are not controversial, e.g., measurement should include multiple stakeholders and should be done at different points in time.

Compulsory Revision: If I understand the diagram, it encompasses the adoption/implementation/sustainment processes. I don’t see implementation outcomes represented. An explanatory note would be helpful in linking this to the bottom line – was the intervention delivered as intended and with what penetration rate.

Issue 4. It is certainly true that the proliferation of home-grown instruments hampers comparisons across studies and efforts to build a cumulative science. The authors point to several reasons that researchers develop their own instruments: the lack of expertise and/or time, the lack of access to established instruments (e.g., the difficulty finding them through literature searches), and the proliferation of terminology in the implementation field. In my experience, one main reason for developing home-grown instruments is the lack of tools with the specificity needed in a particular implementation project. The instruments the authors hold up as exemplary are all generic instruments that do not address issues related to the specific intervention being implemented.

Compulsory Revision: Discuss the need for specificity in implementation measures. If the authors disagree with the need for specificity, counter this with an argument that you can measure all the critical implementation processes and outcomes with generic measures. If the authors agree that specificity is needed to measure some constructs, give exemplars of this type of implementation measure.

Issue 5. Contrasting self-report with mixed methods is a false dichotomy. A parallel contrast would be between self-report and independent observation. Another parallel distinction would be between qualitative and quantitative methods. I think that most researchers endorse mixed methods, but that begs the question of which quantitative methods to use. Throughout the paper the authors focus primarily focused self-report measures. The authors state that “self-report makes good sense” in implementation research. I would counter that assertion by noting that self-report measures typically correlate best with other self-report measures and have a poor track record predicting objective outcomes (i.e., “practical outcomes”). Generally speaking, attitude measures do not predict behavior.

Discretionary Revision: The paper would be much stronger if it were more balanced with examples representing multiple observational methods.

Compulsory Revision: Discuss the value of multiple methods of observations and defend more persuasively the primacy given in this paper to self-report. Fully
discuss the limitations of self-report measurement and how these can be overcome.

Issue 6: The authors interpret a preference for practical measures to mean that respondents want brief measures. While this may be part of it, I suspect that another motivation concerns the goal of finding indicators that can guide programs toward better implementation.

Compulsory Revision: Discuss the interpretation that practicality refers to the desire to have simple quality improvement measures that promote better clinical outcomes.

Issue 7: No concerns.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests