Reviewer’s report

Title: Instrumentation Issues in Implementation Science

Version: 1 Date: 5 December 2013

Reviewer: Gary Bond

Reviewer’s report:

The authors are to be commended for tackling an important albeit severely neglected issue in implementation research – measurement. Many of the issues they raise resonate with my own experience as well as the literature. The identified problems include the lack of common vocabulary for implementation constructs, the lack of shared conceptual frameworks (instead, there is an overwhelming number of frameworks!), the reliance on ad hoc measures without documented reliability and validity, the overreliance on self-report measures, the absence of consistent measure strategies from study to study with the lack of accumulative science. All of these are well-reasoned and sensible points. Most of these criticisms of the field have been made elsewhere, though this paper could well be the most comprehensive analyses of shortcomings in the implementation measurement literature.

1. (compulsory) An implicit assumption of this paper appears to be that the implementation field should have a single unified theory or conceptual framework and that researchers should share a core battery of implementation measures. The hypothesis that a single unified conceptual framework is appropriate across a wide range of practices was expressed in the Fixsen monograph (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005) and is probably a widely shared view among implementation researchers, but it should be made explicit in this paper. As I have stated it, this is two-part proposition, both that there should be a unified theory and that many of the measures should be standardized and used in common. If the authors agree with this proposition, then they should state it explicitly at the beginning of the paper. If I have misperceived their underlying assumptions, then the authors should make explicit their rationale justifying their enterprise of collecting a repository of implementation measures.

The authors endorse the CFIR, the Proctor implementation outcome framework, and the Powell implementation strategy framework as exemplary and complementary frameworks. My colleagues and I also hold these frameworks in high esteem and the strong endorsement in this paper may contribute to what appears to be a move toward consensus in the field, which would be a good thing.

2. (discretionary) There is little doubt in my mind that a number of implementation factors are common across most if not all implementation efforts. To take a simplistic example, adequate funding is almost universally acknowledged as a critical factor sustaining any practice. But is it really true that a common set of factors transcend all implementation projects? In our work, we distinguish among
organizational-level, team-level, and practitioner-level interventions, each of which poses a somewhat different set of implementation challenges (McGovern, McHugo, Drake, Bond, & Merrens, 2013). The authors might clarify the range of interventions for which their measurement repository is intended, or if in fact it aims at a subset of the interventions just described.

3. (discretionary) Even if one assumes that a common theory or framework is appropriate for all implementation projects, there is a second question about the wisdom of using shared standardized measures across studies. Interestingly, one domain that the authors exclude from their project to collect implementation measures is the fidelity domain. They have made this decision because fidelity scales are “necessarily intervention-specific…” (p. 24). But if fidelity measures are specific to the intervention (which is clearly true), then why aren’t other implementation factors also specific? Are the attributes of good supervision identical across interventions? Is acceptability the same? Is penetration? I don’t know the answer, but I do believe that much is sacrificed by reaching for generic measures that by definition asks general questions rather than intervention-specific questions. I also note that fidelity scales are the backbone of implementation research, that in my estimation, most important implementation studies have included fidelity measurement, and the absence of fidelity measurement in earlier research was a huge barrier to building a cumulative science (Brekke, 1988). Furthermore, many fidelity scales are exemplary in their grounding in psychometric, wide scale adoption, and standardization – i.e., some of the very problems the authors identify in the implementation measurement literature. By eliminating fidelity scales from their domain of consideration, they eliminate some of the best instruments promulgated in this field.

4. (discretionary) In my estimation, generic implementation measures, including many very well-accepted ones, often ask such general questions that their precision and specificity is compromised. I worry about encouraging researchers to use generic instruments without modification; I would much rather see research tailor their instruments to their research questions if the study requires it. Certainly researchers often err in the opposite direction of mindlessly adopting instruments that supposedly measure some construct of interest but are far-fetched for their particular application (the so-called nominal fallacy). I did not see this set of issues discussed in the paper and recommend that the authors address this.

Turning now to specific criticisms of the paper:

5. (compulsory) The format of the paper is unconventional. It does not follow the usual structure of Introduction, Methods, Findings, and Results. As a conceptual paper I understand that it does not conform to a research article, but I recommend a more fluid orientation to the paper. The term “debate” introduced in the third paragraph comes out of the clear blue sky. Which debate are the authors referring to?

6. (compulsory) The informal use of survey data from the SIRC conference is also unconventional and frustrating. It is ironic in a paper criticizing the lack of conceptual frameworks and use of unstandardized measures that the authors present a survey that is not formally introduced through a research study. I was
left with the impression that this was a “homegrown” survey instrument (to use the authors' label). It is possible that I am mistaken, but if so, the authors should make the derivation of this survey instrument clear in their paper. The usual methodological questions raised about research quality should also be posed about their survey – for example, how representative is the survey sample? Perhaps the most important question regards the choice of survey questions and what questions were not asked in the survey. How did the authors guard against respondent bias? Did the survey include distractor items? How many of the survey respondents were aware of the SIRC project to collect measures? How many attended workshops given by the study authors? It is relatively easy to get the answers you want by posing questions in a certain way.

7. (compulsory) On Page 17, the authors note that their survey found that “…48% of the reviewed constructs are considered to be in high need of instruments…” This finding is very difficult to interpret without more information. The authors do not present a list of “reviewed constructs” and do not explain what question(s) this statistic is based on. Another example of an opaque reference to survey results is found on Page 21: “…practical instrumentation was indeed prioritized over and above psychometrics with a ratio of 52:48…” There are other examples in the paper. In general, selective reporting of survey results is unacceptable for a scholarly paper. One solution is for the authors to publish the results of their survey study separately and cite that report in subsequent conceptual papers.

8. (discretionary) The authors appropriately criticize the overreliance on self-report instruments. What is not clear is what kinds of quantitative measures the authors endorse. Outside of fidelity scales, most widely-used standardized instruments with established psychometric properties are self-report instruments. I was uncertain which instruments or kinds of instruments the authors were holding up as best practices. The paper would greatly benefit by providing positive exemplars in addition to all the admonitions about the shortcomings of implementation measurement. Tell me how to do it right, don’t just hammer me for what I do wrong.

9. (discretionary) Figure 2 and elsewhere in the paper the authors introduce a substantial amount of terminology, none of which is formally defined in the paper. The inclusion of many undefined terms appears to be incongruent with the authors' criticism of a lack of definitional foundation in much of the implementation literature.

References


**Level of interest:** An article of importance in its field

**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.