Author's response to reviews

Title: Instrumentation Issues in Implementation Science

Authors:

Ruben G Martinez (martinezrg@mymail.vcu.edu)
Cara C Lewis (lewiscc@indiana.edu)
Bryan J Weiner (weiner@email.unc.edu)

Version: 3
Date: 6 June 2014

Author's response to reviews: see over
Dear Dr. Sales and Reviewers,

My colleagues and I are submitting a revised manuscript that seeks to address the editor’s and reviewers’ comments from the second review. We would like to thank you for the opportunity to conduct a second revision of our manuscript. We believe that your expertise and insight into these matters through the review process have greatly improved the quality of the manuscript with respect to both structure and content, as well as increasing its relevance for readers of Implementation Science.

The revised manuscript is sent with this document, which highlights our responses and comments made by the reviewers. Track changes are not presented throughout the manuscript given the great extent of changes made. Instead, we carefully respond to each of the numbered comments below detailing our approach to the revision and point to the section of the manuscript in which the change was made (wherever possible).

Thank you again for the opportunity.

Editor comments

1. What I think is most important is to recognize the need to put some limiters on your discussion and acknowledge that the motivators for your debate paper came from a quite specific and very self-selected group of respondents. In particular, I note that the two listservs you used to identify implementation researchers (and perhaps that would be a preferable term to "stakeholders", which often has specific connotations in implementation research) were not fully representative of the field. You do indicate this, but you don't state this explicitly as a limitation in your discussion. I think this is quite an important limitation-- the issues affecting researchers in the areas you surveyed are to some extent overlapping but also distinct from implementation researchers in other areas of health care, which is the main scope and focus of the journal. I think you can, with some additional editing and clarification, ensure that this somewhat skewed perspective may have influenced your recommendation in ways that may not have been so pronounced had you surveyed a different group of researchers.

We have modified our language to more accurately describe the field of expertise of our respondents, address the limitations of the survey and the relevance of our findings, and have explicitly stated the potential limitations of this perspective.

2. As we usually require, please ensure that you provide a point by point response to the review comments, and please ensure that the final manuscript is within the 5000 word limit.
Reviewer #1:

3. 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.

We appreciate your experienced, critical viewpoint and believe it will serve to strengthen the foundation of the manuscript. As stated in our response to the editor’s comment #1, in this second revision, we have made clear that the perspectives presented in this paper are reflective of a subgroup of researchers engaged in implementation science, which may differ from others engaged in implementation science. We have included this as a limitation in the manuscript. Simultaneously, we attempt to present a more balanced presentation of measurement options in the recommendations section of the paper. Moreover, we hope that the Debate format encourages dialogue in the field to address instrumentation issues from a variety of perspectives to determine the degree of overlap.

4. 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.

We have included a section on limitations (e.g., potential for biased responding given modest response rate) of our sample (response rate) and the survey in the background section of the manuscript to ensure that readers have an appreciation of the perspectives represented from the outset.

5. Issue 1: Looks fine.

Minor Essential Revision: Please explain what a Wiki is.

We have added a sentence that defines the term “Wiki”.

6. 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.

We completely agree with this point and have edited the Psychometric section to explicitly state the need for appropriate reliability assessment depending on the instrument method.

7. 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?

Due to space limitations, the section on the EBPAS has been removed from the manuscript.

8. 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.

It is helpful to learn your interpretation of the diagram. However, due to space limitations we chose to remove the “what to measure, when, and at what level” section. Thus, the diagram is no longer included in the manuscript. This decision was made also because this section has raised concerns and confusion at each stage of the review and it was not an issue that was originally included in the survey.

9. 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.

This is a tricky issue because a cumulative science could be most efficiently built on generic instruments, yet simultaneously it seems logical and reasonable to expect that specific instruments may be necessary to evaluate a particular implementation. Because we have
been unable to locate literature that supports the claim that specific instruments are necessary, we strongly urge the field to consider use of generic instruments. As such, we acknowledge the potential need for both types of instruments and highlight the importance of additional research to resolve this issue and guide the field.

10. 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.

We have altered the title of the section to more accurately portray that we do not wish to contrast the two approaches; rather, we our goal is to examine the strengths and limitations of both methods and approaches while attempting to advocate for more deliberate selection of one evaluation method or approach over another.

11. 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.

12. 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.

This point is well taken and we have integrated this reviewer’s language into the revised manuscript, with an attempt to balance the discussion of self-report with multiple methods of observation. We also carefully discuss the limitations of self-report measurement and how these can be overcome.

13. 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.

We have added a sentence that addresses how our respondents viewed practicality and the limitations of this response to better conceptualize what was meant by “practicality”.

Issue 7: No concerns.

Reviewer #2:
The authors have done an excellent job responding to suggestions for the initial submission. They have provided a number of very good recommendations, which should be useful to implementation science researchers. I have a few new major compulsory and discretionary revisions to request.

Thank you. We believe that the concerns you put forth initially have been critical in allowing us to strengthen the manuscript.

**Major Compulsory Revisions**

1. Please define the term “stakeholders” at the start of the manuscript. Whom does this include? The way the term is used throughout most of the paper it seems that it refers primarily to implementation researchers; but the term often has a broader meaning than this, so it would be good to clarify.

We have decided to replace the term “stakeholders” with researchers/respondents (as appropriate) given the typical connotation for stakeholder would not include researchers, whom were the primary perspective reflected in this paper.

15. I don’t think that the authors have adequately addressed one of my original concerns (also mentioned by Reviewer #1)—namely, that insufficient information is provided on how the survey was developed. The manuscript states, “Instrumentation issues were generated by the authors based upon lessons learned through the Seattle Implementation Research Collaborative (SIRC) Instrument Review project. Items were refined by SIRC’s core research team members.” It would be helpful to have a little more information on what activities/information generated the “lessons learned.” Based on how it was developed, would the authors suggest that this is a fairly comprehensive list of issues? Or would they suggest that the issues are unique to the SIRC work? For example, I’m curious about how the issue of “over-reliance on self-reports” was determined to be an important item to include in the survey. (See my comments below on Instrumentation issue #5.)

In our effort to improve clarity of the manuscript content and reduce focus on the SIRC Instrument Review Project, we compromised the ability to explain the genesis of the issues identified. Although the issues identified may be most pertinent to SIRC’s activities, this work is relevant to the evaluation of constructs delineated both in the Consolidated Framework for Implementation Research (Damschroder et al., 2009) and the Implementation Outcomes Typology (Proctor et al., 2009). These two frameworks are among the most highly cited in the broad field of implementation science. As such, we hope that the issues identified through the SIRC search and synthesis project reflect broad issues worthy of the field’s attention. Indeed, we have tried to frame these issues as broad-reaching instrumentation issues for the field of implementation science. To avoid reverting back to the inclusion of the SIRC work, we strive now to ground the issues in the published literature. Moreover, the issue of “over-reliance on self-reports” has been renamed/reframed.
16. Page 20, Step six: Creating a report. Please clarify what is meant by “factor/subscale information” and “normative data.”

We have slightly altered this sentence to point to information in “Issue 2” and provided an explanatory note regarding normative data.

17. Pages 20-23, Instrumentation issue #5: Overreliance on self-reports and underuse of mixed methods. I didn’t mention this in my first critique; but on reflection, I’m concerned that the problem with self-report may be overstated here. To the extent that self-reports present peoples’ perceptions of something, and perceptions’ play a huge role in implementation, I’m not sure what “over-reliance” refers to. It’s not like we’re trying to measure an objective construct like physical activity or diet, which can be significantly distorted via self-report. We want to know how difficult an implementation is perceived, or whether people perceive there to be sufficient evidence for an intervention—and for these, it seems like self-report is the best way to obtain this information. This is acknowledged in the statement, “Use of self-report makes good sense given that many of the salient constructs pertain to individuals involved in implementation. For instance, researchers have discovered through the use of self-report that individual knowledge, beliefs, and attitudes toward an evidence-based practice can serve as a barrier to implementation.” And certainly it makes sense that fidelity is a construct that should be measured more objectively (perhaps by observation, as suggested). But this discussion, and the title of the section, suggests that there’s too much selfreporting— how then do we obtain information on barriers and facilitators to implementation, which is a key component of implementation research?

Both reviewer 1 and 2 make excellent points on this section that has led to major revisions (see Reviewer 1 comment #3). We hope that this revision addresses both reviewer’s concerns.

18. In addition, I’m not sure it’s appropriate to combine the issues of self-report and mixed methods. If these are going to be combined, I think the terminology needs to be a little clearer. I think selfreport would be considered a source of data—as observations and administrative data bases are also sources. Surveys, interviews, and focus groups are all different methods for obtaining selfreported data. Qualitative vs. quantitative approaches can encompass both data collection and data analysis. Including the issue of self-report with the discussion of mixed methods muddies the terminology a bit. For example, the last paragraph of this section includes the following statement: “In sum, self-report is important for assessing many key implementation constructs, but the limitations of relying strictly on self-report should be carefully considered.” This statement follows a discussion of mixed methods—it is a non-sequitur. The next statement is: “Additional options for diversifying assessment methods include mixed methods, multi-informant approaches, direct observation, as well as the use of administrative and existing data such as those captured within the soon to be ubiquitous electronic health records.” Again, this seems to mix up data collection and analysis methods (mixed methods) with data sources.

This concern was also noted by reviewer 1 and we have, as stated above, greatly revised this section of the manuscript to clarify this issue.
19. Page 27, under Summary. “Ultimately, through this article, implementation scientists might be better able to determine what type of evaluation is most appropriate for which interventions and settings…” What is meant by “type of evaluation”? Plus it’s not clear to me that this paper provides sufficient detail for identifying a particular evaluation approach based on intervention and setting. If the authors feel that this is true, some additional explanation for how to do this is needed.

**In response to this statement, we have edited the statement to read “…Implementation scientists might be more equipped to think critically about instrument development and administration, the factors influencing the quality of instrumentation, and the limitations and strengths of different instrumentation and evaluation approaches.”**

**Discretionary**

20. Page 6. “However, no literature, to our knowledge, addresses the convergence/divergence of distinct construct definitions across models.” As a framework, the CFIR does address this issue: “We combined constructs across published theories that had different labels but were redundant or overlapping in definition, and we parsed apart constructs that conflated underlying concepts” (from the abstract in Damschroder LJ, Aaron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science* 2009;4:50).

However, the CFIR paper does not explicitly state which constructs from other theories map to which CFIR constructs (or vice versa), which is perhaps what the statement is referring to. Perhaps this can be clarified?

**While the CFIR does combine and parse apart different constructs, it does not convey how construct definitions differ across models (i.e. how is “adoption” defined in Greenhalgh’s model as opposed to Proctor’s as opposed to NIRN’s?). We have revised this sentence to more carefully explain our meaning.**

21. Page 7. Add the word “terminology” to: “Despite the rapid growth of our field, it remains unclear what factors are critical for successful, sustainable implementation, in part because of inadequate and inconsistent use of theory, terminology, and measurement.”

**We have added “terminology” to this sentence.**

22. Page 7. Suggest adding “including providing specific examples of how constructs are operationalized in various studies” to the end of the statement, “This Wiki is a coordinated effort that aims to address this issue through an online collaborative website that allows stakeholders to establish and refine implementation-specific terms and conditions.” It is through these specific examples that we hope to define constructs in a clearer and more consistent manner across studies.

**We have added your suggested wording to this sentence.**

23. Page 8. “In addition, stakeholders looking to develop instruments can more readily base their items on relevant theory, a process that is integral to the psychometric validation of an
instrument.” This statement is a little unclear as it pertains to use of the CFIR. The CFIR is not a theory—i.e., it does not depict interrelationships among constructs or present specific hypotheses of which constructs are the most important ones for affecting implementation success. Instead it includes constructs from a synthesis of existing theories. Its contribution is in specifying the terminology and definitions of constructs from these various theories. Therefore, I would suggest deleting this statement.

We have deleted this statement from the manuscript.

24. Page 8. The CFIR actually does include constructs from the Diffusion of Innovations Theory. It’s subsumed within Greenhalgh’s framework. Related to the previous comment about CFIR and theory, I would suggest revising the 2nd paragraph to better explain that the CFIR is not a theory, and that readers may want to look to May for this. I might suggest revising the paragraph as follows to address these two issues:

“Although the CFIR is relatively comprehensive with respect to identification of salient implementation constructs, the framework does not include implementation outcome constructs, nor does it directly overlap with other widely used models. For instance, survey respondents reported frequent use of the Active Implementation Framework [20] (3rd most frequently endorsed), which outlines constructs unique from those included in the CFIR. Although there may be benefit to endorsing a single conceptual model for use in implementation science (enables cross study comparison, facilitates identification of predictors, moderators, and mediators), there are also inherent disadvantages to settling on a unifying theory early in a field’s development (limits discovery, overlooks understudied constructs). In addition, The CFIR is not a theory—i.e., it does not hypothesize interrelationships among constructs. For a comprehensive theory of implementation, readers may wish to consider the general theory of implementation proposed recently by May [21]. At a minimum, implementation researchers are encouraged to include construct definitions in their study reports to promote transparency of their work and generalizability of their findings.”

We have revised the paragraph per your suggestion.

25. Page 19, Step two: Initial item development. “Researchers may begin by engaging think-aloud techniques as a possible method for developing instrument items.” I think the think-aloud technique is better used for testing items that have already been developed—at least, that’s how I’ve seen it used most often. I.e., using it when administering a survey to respondents and asking them to think aloud as they respond to each item, describing how they interpret the item, why they’re answering it the way they are, etc. I’m less clear about how it’s used in item development. I would recommend moving it to step three, substituting the word “testing” for “developing.”

We have removed the think-aloud section from the development section and added it to section regarding pilot testing.