Author’s response to reviews

Title: Developing measures to assess constructs from the Inner Setting domain of the Consolidated Framework for Implementation Research.

Authors:
Maria Fernandez (maria.e.fernandez@uth.tmc.edu)
Timothy Walker (Timothy.J.Walker@uth.tmc.edu)
Bryan Weiner (bjweiner@uw.edu)
William Calo (wcalo@phs.psu.edu)
Shuting Liang (Shuting.Liang@emory.edu)
Betsy Risendal (Betsy.risendal@ucdenver.edu)
Daniela Friedman (dbfriedman@sc.edu)
Shin-Ping Tu (sptu@vcu.edu)
Rebecca Williams (rebeccawilliams@unc.edu)
Sara Jacobs (sjacobs@rti.org)
Alison Herrmann (aherrmann@ucla.edu)
Michelle Kegler (mkegler@emory.edu)

Version: 1 Date: 05 Jan 2018

Author’s response to reviews:

Dear Dr. Squires,

We are pleased the reviewers thought our manuscript was “very interesting” and reflected “tremendous work in an important area.” We thank you and the reviewers for considering our manuscript. We have addressed the reviewer comments as detailed below.
Reviewer #1: This is a very interesting paper that logically proceeds from an existing framework (CFIR) to develop measures of the inner setting. The methods appear to be relatively straightforward and carefully done. The authors are also forthcoming in their limitations section (given the underlying philosophical assumptions).

1. "we chose CFIR constructs that were relevant for FQHCs, modifiable, and could be measured with few items." How could the authors be certain that a few items truly captured the constructs (both construct and content validity)?

Response: We agree with the reviewer that we cannot be certain about the number of items needed to capture a construct. We updated the text on page 8 in the first paragraph of the Development of Measures for Inner Setting Constructs section to clarify that we chose constructs that we hypothesized to be measurable with a few items.

Given the purpose of the paper was to determine whether the items selected were sufficient in assessing each construct, our methods assess both construct and content validity of the items chosen during our measurement development process.

2. One of the criteria for item selection was that the "demonstrated some evidence of reliability and validity in previous research." How much and what type(s) were necessary to meet this threshold?

Response: We updated the text on page 9 in the Identification and Selection of Items section to clarify our item selection process. Specifically, we now state that items were selected from published studies with measures that demonstrated some evidence of reliability (e.g., internal consistency) and/or validity (e.g., face validity, construct validity) in previous research.
3. A quibble: is there a difference between EBA and EBP? Do we need a new acronym?

Response: We decided to use and maintain the term EBA as we feel it broadly captures all types of interventions, programs, guidelines, etc.

4. It is unclear what is mean by "highest expected prevalence" in the statement the EBA inserted into the item was the one with the highest expected prevalence …

Response: On page 11 in the Identification and Selection of Items section, we updated the language to better explain how an EBA was inserted into a survey item. We now state that follow-up questions were based on responses about which EBA they were implementing. If they were implementing more than 1 EBA, the follow-up questions inserted the EBA prioritized in the following order: provider reminders first, followed by patient reminders, one-on-one education, and provider assessment and feedback.

5. Response rates: Although it is possible to figure out some of the response rates, e.g., clinics, it is not at all clear what the response rates are for the different health care providers and whether there was differential response from different clinics. Were non-responding clinics similar to responding clinics?

Response: On page 15 in the Sample characteristics section, we added information about the average number of respondents per clinic and the breakdown of the number of responses per clinic. Given the data collection and recruitment approach, we do not have information about the number of people eligible or approached per clinic, and therefore, cannot provide exact rates. We do not have information about non responding clinics for comparison.
6. The authors identified two subconstructs - culture effort and culture stress. What is gained by separating these out from the other subconstructs in the culture domain? I am also not convinced that there isn't overlap in some of the items in different constructs, e.g., items A40 and A7 and A22.

Response: Based on our literature review of existing measures and constructs, we feel culture stress and culture effort are important constructs to be able to measure. This is because they are theoretically distinct from general culture. Additionally, effort and stress have been shown to be independently associated with productivity, and working environment, which could both impact implementation outcomes.1-2

We would expect there to be some overlap between items given they are capturing an element of culture. This is especially true for A07 and A22 since they are items capturing general culture. We also would expect A40 to be related to these items because culture effort is related to general culture. Nevertheless, our results demonstrate that culture effort and general culture, while correlated, are distinct constructs.


Reviewer #2: Dear Maria E. Fernandez and team,
Your article reflects tremendous work in an important area, and I think it will be of great value to the field. My main question related to your article is: Are you trying to achieve too much in a single article? Viewed from my perspective, what you have submitted lends itself to two articles:

1) One describing the development of the measure

2) One describing the validation of the measure

Response: As recommended by the editor, we are keeping the manuscript as one paper covering the development and testing of the measures. Nevertheless, we have attempted to respond to this reviewer’s other suggestions.

In attempting both of the above in one article, you have to limit your word count and that leads the section on the development of the measure to fall short in answering questions I consider important to answer so readers have a chance to fully understand what decisions and selections have been made to develop it. The most central question would be:

1) What exactly did you do - in overarching terms. If my reading of the article is correct, you
   a) Selected the inner context CFIR domains - there were five of them

   b) You found measures - by searching for surveys that were already used in your system and by looking for measures in the literature - that met certain pre-defined criteria
c) You then selected items from these measures that potentially could fit the construct (?) - and this is where you used a consensus development process? This is a step I am most uncertain about because it is not clearly described.

d) You then also seem to have worked the other way round: You used measures to expand on the CFIT constructs because you came across items in the measures you identified (stress / effort) that appeared to be helpful in this regard (?) - again, this is something I am uncertain about because it is not fully described how these decisions were made and why.

Response: We updated the text in the Development of Measures for Inner Settings Constructs section starting on page 7. We clarified the consensus development process and how we selected the inner setting CFIR domains, how we identified measures, and how items were selected.

The reviewer makes a good point that in one case, we expanded the CFIR construct (culture). We clarify this in the description of our approach on page 9 to highlight the iterative nature of the development process.

Furthermore, and linked to the above, I would have additional questions that I think would need to be addressed more clearly in an article about the development of a measure. Among these are:

1) What have been the steps taken in what is described as the 'consensus development process'? Who was involved, what were decision making guidance and procedures etc.? In its current form, the section describing this part of your work appears not fully transparent to me, which - just as an example - makes it difficult to understand how you landed on 15 out of 37 CFIR constructs.
Response: We have provided more information about the consensus development process in the Development of Measures for Inner Settings Constructs on page 8.

2) In identifying items for the survey you describe that you did draw on existing surveys and especially on the AAPCHO - what other surveys were included and why the choice pro AAPCHO? 3) What were the search terms used for identifying literature on relevant measures? How were these reviewed and selected? I understand that you did not conduct a systematic review of the literature but that makes it even more important to describe the steps taken in this part of your process.

Response: In the Identification and Selection of Items section on page 8, we added information stating that we chose AAPCHO because it was highly related and allowed us to build on previous work.

We also added examples of the search terms when reviewing literature for measures and constructs and information about reviewing the Grid Enabled Measures (GEM) and Society of Implementation Research Collaboration (SIRC) websites.

4) From the measures identified as relevant to your work: How did you decide on the selection of items / questions under each construct? What characteristics guided your decision making?

Response: In the Identification and Selection of Items section on page 9, we clarified that the information presented was the criteria used for the item selection process. We also provided more information about what type of evidence we used to inform inclusion as also suggested by reviewer 1.
5) What lies behind the process of seeking input from leaders to ensure the appropriateness of the measures? How did you retrieve that feedback? What has the feedback been? Did it imply for you to make any changes?

Response: In the Item Adaptation and Survey Development section on page 11, we provided more information about what we specifically asked the leaders to review and the type of feedback we received.

By developing the measure development into a separate article, you will be able to provide details like the above and create a better understanding of how choices and selections were made. I would then also include a presentation of the full measure - which I missed in the current version of the publication. It would be helpful to see the entire list of statements respondents were asked to rate together with the rating scale used.

Response: As stated above we are following the editor’s recommendation to keep this as one paper. We have amended Table 3 to include the response categories so it represents the full measure.