Reviewer’s report

Title: A Mixed-Methods Study of System-Level Sustainability of an Evidence-Based Practice in 12 Large-Scale Implementation Initiatives

Version: 0 Date: 04 Mar 2017

Reviewer: Gary Bond

Reviewer’s report:

This paper tackles an important but difficult issue - sustainment of an evidence-based practice (PCIT) in across a state (system). The authors examined 12 different systems and used both qualitative and quantitative methods. In their quantitative methods they used several scales. The study represents a huge level of effort.

I had a number of concerns of this paper starting with the clarity and precision of their literature review.

In the Abstract, the authors state that "evidence-based practices are becoming more widespread..." Please give a citation for this. Based on some longitudinal surveys, it's actually not clear if evidence-based practices are becoming more widespread (Bruns et al., 2016). The Abstract also uses the phrase, "sustain (facilitate and optimize)," which is unclear. I did not see in the paper where optimize is defined. I would also modify the Abstract so that the Conclusions do not simply repeat the Results.

In the Introduction and Discussion the authors cite the Stirman review paper (2012) on sustainability repeatedly, which is very appropriate, given that it is the most comprehensive and most recent review of the sustainability literature. The Stirman review notes that 65% of the studies they reviewed did not give an operational definition of sustainability, which is one of the fundamental problems in this research area. In the first paragraph of the current paper, the authors state that sustainability "is generally thought to be the process of maintaining or improving a system's ability to preserve a program's function and utility under continued change," which I do not find to be clear. Given the confusion in the field, I believe it is incumbent that the authors provide a crisp and understandable definition.

Also in the first paragraph, the authors state that "Sustainability typically occurs two or more years after implementation." I think that the time frame issue warrants more discussion. It is certainly a vexing issue in this area, but what the authors should convey is the arbitrariness of the time frame in many sustainability studies. The Stirman review delves into this issue and helps clarify differences in how different research groups have defined the so-called sustainability phase. The sentence in question goes on to say that systems attempt to optimize the fit during this period. Does this mean that before this period, systems are not attempting to optimize? What does this sentence mean? Throughout the Introduction the writing style is overly abstract and difficult to understand.
The second paragraph of the Introduction distinguishes between "smaller scale implementation" and "large-scale initiatives." As an example of the former, the authors cite Swain et al. (2010) and as one example of the latter, the authors cite Bond et al. (2014). The authors may not realize this, but these two papers are based on the same study (National EBP Project). It is true - and worth noting - that system-level influences are present in both small-scale and large-scale projects. If that is the authors' intent, then that is what the report should say.

Moving on to the Methods, the selection of the 12 initiatives is well described and appropriate. The one concern I had was that the initiatives ranged widely in time of existence, from 3 to 23 years. It does seem to me that this is an important confound. After 23 years, you may still be training new clinicians with inevitable turnover, but the rate may be lower than for a more recently-established initiative. I did not see years in existence included in any of the statistical analyses. The report should address whether this variable influenced sustainment or any of the auxiliary measures.

The variability of the stakeholder groups is not ideal and represents a significant study limitation. The authors should note this. The authors do not explain why they do not have interviews with all 12 (or 13) state leaders, who surely have a pivotal role in the systemwide implement and sustainment of PCIT. Please also explain the snowball sampling approach - I suppose they were seeking the best informants, but would a judge have similar information as an academic? The sampling strategy seems haphazard. It's quite possible that there are good reasons the authors chose to proceed in this fashion. Please explain.

The authors were apparently able to obtain a 25-item survey from all primary trainers. Why, then, did they not ensure that all these trainers complete the sustainment surveys? This would give a specific stakeholder perspective. This would have been cleaner than obtaining "key stakeholder" surveys to represent the viewpoint of the initiative.

On Page 8 the authors refer to "an overall rating of sustainability." Is it correct that the raters (i.e., the interviewer and the key informant) made this rating without having any standard operational definition of sustainability? The authors should discuss this.

The description of the qualitative analysis in the Methods seems fine. The authors report the range for the kappas. As they are of course aware, a kappa of .29 is unacceptable. The range does not tell us how many kappas were unacceptable, but the authors need to comment on this and report this as a limitation. Implicitly they are saying, don't look at the kappas, we reached consensus after discussion. The Results section includes numerous quotations from the qualitative interview, which I found to be excellent and a strength of the paper. I did not see anywhere that the Atlas coding was actually incorporated into the report except through the quotations. As such, the claim of a mixed methods approach is weak.

The biggest methodological weakness in the paper concerns the measurement of barriers, strategies, and the PSAT. The authors obtained ratings from two perspectives - interviewer and "initiative" (the key stakeholder) to two surveys comprised of 19 items and 40 items, respectively, where respondents selected from a 7-point semantic differential scale to various terms (e.g., "Openness to EBPs"). I assume that the terms were not defined. The authors refer to
ratings on the semantic differential scale, ranging from 1 "to a little extent" to "to a great extent," as objective ratings. In what sense are these objective ratings? They are quantitative ratings, but clearly they are subjective opinions, filtered through the respondent understanding of the terms on the checklist. The ratings are not behaviorally anchored. There is a large literature on rater bias, which this paper does not mention.

The Results section reports the descriptive findings from the survey instruments using terms such as "high," "mid-to-high" and "low," based on the ratings. One troubling aspect of the report is that interviewer and initiative ratings do not seem to agree very much. The authors should analyze the agreement statistically (possibly using a kappa, or another statistical measure). Assuming that my impression is correct, that there is large disagreement, what then are we to make of this? The authors do discuss various interpretations - different information, different time frame, etc. That being said, what are we as readers to believe is in fact the best evidence?

In addition to these measurement issues, I found the presentation of these main results tedious and difficult to follow. One alternative method of presentation would be to take head on the problem of the lack of agreement between the two sources (interviewer and initiative) and indicate that the research team concluded that the initiative perspective was more credible and therefore the paper presents all the findings in terms of that perspective only. The comparison between sources could be addressed in a separate section and separate table. I don't see how the interviewer perspective is really helpful. I do see that the authors have chosen the initiative rating of overall sustainability in their multiple regression, so they are implicitly agreeing with me that the initiative data are probably more credible.

Please possible typo on Page 15 in the paragraph under subheader, "Balancing supply": I believe the first ratings are interviewer ratings.

Given uncertainty about the validity of the ratings, I am less confident about the multiple regression analyses to predict sustainability outcomes. While the authors do mention Bonferroni corrections, they do not address the small sample size and resultant lack of statistical power.

In the Discussion, the authors cite a sustainment rate of 89% from the Stirman review. I have not located that statistic in the Stirman review.

The Study Limitations section needs to be more comprehensive.

References


**Level of interest**
Please indicate how interesting you found the manuscript:

An article of importance in its field

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable

**Declaration of competing interests**
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report
including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal.