Author's response to reviews

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

Authors:

Ashley Scudder (abtscudder@gmail.com;scudderab@upmc.edu)
Sarah Taber-Thomas (taber@buffalo.edu)
Kristen Schaffner (schaffnerkm2@upmc.edu)
Joy Pemberton (JRPemberton@uams.edu)
Leah Hunter (ljh222@psu.edu)
Amy Herschell (Amy.Herschell@mail.wvu.edu)

Version: 1 Date: 26 Jun 2017

Author’s response to reviews:

June 26, 2017

To the reviewers:

Thank you for reviewing our manuscript [Manuscript No.: HRPS-D-00166] A Mixed-Methods Study of System-Level Sustainability of Evidence-Based Practices in 12 Large-Scale Implementation Initiatives] for publication in Health Research Policy and Systems. We again are extremely appreciative of the time and energy you have devoted to this manuscript. We have revised this submission to incorporate the provided recommendations and have adhered to the 5500 word limit. Below we have outlined reviewer comments as well as descriptions of our modifications.

1. Reviewer #1: Page 4, line 15:

It would have been interesting to see sustainability not just as a phase but as a continuous action and taking into account all the intervention, not just its "aftermath".

See:https://www.researchgate.net/publication/222928496_Program_Sustainability_Begins_with_First_Events

https://implementationscience.biomedcentral.com/articles/10.1186/1748-5908-8-117
Response: We agree that sustainability involves continuous action, not simply the aftermath of implementation. Thank you as well for the two suggested citations. We have reviewed these, and have included them in our resubmission. To better reflect the continuous action of sustainability in our writing we have removed the use of the word “following” from the title and the text (e.g., Page 4). We agree that this is more consistent with our view of sustainability and particularly why we included so many details related to the initial implementation. We have also added related text to our introduction and discussion sections (Page 20).

2. Reviewer #1 Page 6, line 3:

Was the conference call part of the data collection? if yes, can it be more defined? (who took part, when, was it part of the trainers’ survey or the sustainability phone interviews?)

Response: Yes, the conference call was the method used to interview initiatives regarding the sustainability of PCIT. We have adjusted the text (Page 7) to clarify that these calls were the way in which data was collected for the “sustainability interview.”

3. Reviewer #1 Page 6, line 18:

Are the "1 to 3 calls" attempts to call? or number of participants? (was there not 1 to 8 participants/initiatives?)

Response: Thank you for this clarification. We have adjusted the text to clarify that the number of calls required to complete the sustainability interviews ranged from 1-3 across initiatives, and that the number of participants on the calls ranged from 1-8.

4. Reviewer #1 Page 7, line 35:

Participants are the "initiatives representatives"? or the larger group of 37 participants?

Response: These were individuals within the larger group of 37 participants, who were identified from each initiative to respond. We have changed the wording to clarify this.

5. Reviewer #1 Page 8, 1st paragraph:

I assume the sustainability outcomes were perceptions by the participants, it would therefore be important to specify it. Was there any document used to triangulate the information? if so, it would be interesting to add this in the methodology/results.

Response: Both subjective (sustainability rating) and objective data (% clinicians, % agencies continuing to provide PCIT) were included as sustainability outcomes. Objective data
was pulled from training or state records. Though it was highlighted in Table 2, the text was adjusted to highlight these results as well.

6. Reviewer #1 Page 9, line 18:

Was there a specific definition of sustainment and its different levels? or was it based again on personal conceptions?


Response: In order to be eligible for the study, initiatives had to be past the initial 12-month formal training and implementation period. This broad definition of sustainability was provided (i.e., continuation of PCIT beyond training period), however, due to the exploratory nature of our study, the aim was to understand initiatives’ perspectives in order to refine our conceptualization and definition of sustainability. Many initiatives identified sustainability-specific activities that occurred early in the implementation process, and this point was incorporated into the discussion (Page20).

7. Reviewer #1 Page 18, line 65: Is there an explication of why some initiatives reported financial support as a lesser barrier? Was it because they had more financial support than others?

Response: This is an interesting question, and we have included additional text to the discussion section to address this (Page 19).

8. Reviewer #1 General:

- There is no discussion on how the methodology employed could have influenced the results: phone + online interviews, snow-ball data collection. Can there be a possibility of overestimating the results due to these methods (no face-to-face confrontation or sampling bias of the snow-ball methods)? Could this also explain the variations between the interviewers’ and initiatives' ratings?

Response: The reviewer highlights several considerations regarding the potential impact of methodology on the findings. While a snow-ball sampling method could lead to only certain participants selecting into the study, because these initiatives were circumscribed efforts, it was possible to identify nearly all individuals involved using this method. Moreover, 72.5% of individuals reported as key stakeholders agreed to participate. Thus, it seems unlikely that this method would have biased the results. We have attempted to address this point by more clearly describing the methods (Page 7).
9. Reviewer #1 - Again, is there a possibility that participants gave good accounts of the initiatives because it looked like an evaluation? Is there room for a reflexive thought?

Response: The reviewer poses an important question. Though the reliance on retroactive, initiative-reported data may be seen as a limitation in the given study, we worked to address this in the selection of study participants. This is specifically why we did not limit interviews to PCIT trainers alone and felt that it was important to include others involved in the initiative. In addition, the study was described as being an attempt to gain a better understanding of large scale initiatives, not to evaluate their “success.” As such, conference calls were characterized as open discussions in order to learn about the individual initiatives, not as evaluations. Further, we made all possible attempts to de-identify information when possible, so that individual participants felt more comfortable being candid. We encouraged participants to discuss both positive and negative aspects of their efforts, in order to better inform future initiatives about potential challenges and successes of PCIT implementation and sustainability.

10. Reviewer #1 - In the Methodology section, since there are different participants for different data collection processes, to integrate the "participants and procedures" section within the "Measures" section wouldn't make it clearer and less redundant?

Response: Thank you for this suggestion. We have integrated the measures into the participants and procedures section to eliminate redundancies.

11. Reviewer #1 - Since it is a mixed-method study, it would be interesting to integrate more the qualitative with the quantitative data. Eg.: Clearer analysis processes, a rational why mixed-method was used (since the quantitative side is a bit weaker via its few participants), the limitations of this methods, etc.

See:


Response: Thank you for this suggestion. We attempted to better integrate the different types of data throughout the results section, as well as in the discussion. In addition, we included a clearer description and rationale for the data analysis approach used (Page 8-10).

12. Reviewer #1 - A better description of the context(s) of the initiatives would be interesting and would enable more transferability of the study. Many criteria are however there in the study, which is very good (Qualitative approach, reflexivity, etc.).

See:https://www.ncbi.nlm.nih.gov/pubmed/24979285

Response: Thank you for sharing the Standards for Reporting Qualitative Research document. The methods and results have been edited substantially. Due to space limitations, we were limited in how much context could be provided, but attempted to add relevant details where possible, throughout the revised manuscript.
13. Reviewer #1 - Very interesting article that confirms many other studies (that could have been in the discussion), with many aspects that will now have to be further explored. (pdf version of the referred articles are joined to the review)

   o Response: Thank you for your feedback. We agree that the addition of additional citations will strengthen the resubmission, and the discussion section has been substantially revised to incorporate relevant studies in the literature.

14. Reviewer #2: 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.

   o Response: Thank you for your thoughtful review and acknowledgement of the level of effort and time that this project entailed.

15. Reviewer #2 In the Abstract, the authors state that "evidence-based practices are becoming more widespread..." (Page 3, line 16) 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).

   o Response: Thank you for your comment. The abstract has been revised and this wording removed.

16. Reviewer #2: The Abstract also uses the phrase, "sustain (facilitate and optimize)," which is unclear. (Page 3, line 18) I did not see in the paper where optimize is defined.

   o Response: Thank you for your comment. We have worked to tighten this writing in the abstract and throughout the manuscript.

17. Reviewer #2: I would also modify the Abstract so that the Conclusions do not simply repeat the Results.

   o Response: Thank you for your comment. The abstract has been revised accordingly.

18. Reviewer #2 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. (Page 4, line 21). Given the confusion in the field, I believe it is incumbent that the authors provide a crisp and understandable definition.
Response: Thank you for this feedback. The introduction has been substantially revised, and we expanded the review of varying definitions and conceptualizations of sustainability. In addition, one goal of this study was to incorporate initiatives’ perspectives in order to refine our conceptualization and definition of sustainability. Thus, although initiatives were excluded if they had not yet completed the initial training period, we collected information about sustainability-processes occurring over the duration of the initiatives. This has been more explicitly stated in the introduction (Page 4) and methods sections (Page 6).

19. Reviewer #2 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 (Page 4, line 18). 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.

Response: We appreciate this reviewer comment as we feel that it has helped us to tighten our writing and to add clarity to the introduction of the paper. This comment hits on the concept of sustainability as a continuous process versus phase, which we agree with, and feel that the text of our introduction did not reflect. We have edited our text to provide clarity.

20. Reviewer #2 The second paragraph of the Introduction (Page 4, line 28) 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.

Response: To clarify the reviewer’s point we have removed Swain as an example (Page 4). Although this article describes agency-level versus system-level data, it may distract the reader given it was part of a larger examination (National EBP Project).

21. Reviewer #2 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.

Response: Thank you for raising this point. We were initially thinking along similar lines, and prior to the first submission we examined whether the “length” of the initiative was important. Specifically, initiatives were sorted into two subgroups of oldest (beginning in 2004
or earlier) and youngest initiatives (beginning in 2009 or later), and we examined differences in qualitative findings, as well as whether there were differences in outcomes based on time since initiative began. Neither analysis yielded significant findings. Thus, due to space limitations, this was not included in the submission.

22. Reviewer #2 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 implementation and sustainment of PCIT. Please also explain the snowball sampling approach (Page 5, line 62)- 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.

Response: The methods section has been revised to clarify the significance of stakeholder roles to each initiative (Page 7), as well as our use of a snowball sampling approach (Page 7). It was our experience that initiatives identified anyone that they felt played an important role or might contribute to the perspective of the initiative; however, within each initiative there were several key stakeholders who had been intricately involved and knowledgeable of the initiative efforts. During the conference calls, these individuals were often relied on to provide specific details or identify additional individuals who could better answer the questions. When EBP trainers were not one of the most knowledgeable stakeholders, it was most often the case that someone at the state-level or in a governmental role was. Finally, although some initiatives included unique stakeholders (e.g., judge), these individuals typically had access to unique/important information regarding aspects of the initiative that others could not report on. In short, their inclusion allowed us to gather the most complete information.

23. Reviewer #2 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.

Response: While we agree with the reviewer’s point that providing sustainment surveys to all trainers would provide a specific stakeholder perspective, we felt that trainers (who in some initiatives were only involved with projects for a 12-month training and implementation period) would not best be able to address the sustainment across time as would initiative leaders who were actively embedded into state systems. This has been noted in the revised manuscript (Page 8) In addition, it was our belief that training in multiple initiatives might influence some trainer’s perspectives. For those trainers who train across the country, it’s likely that they are not aware of all the state specific nuances that were relevant and that if would have been difficult to keep some details initiatives clear.

24. Reviewer #2 On Page 8, line 11 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.
Response: We appreciate the reviewer’s request for clarification in the text. Only initiatives provided an overall rating of sustainability. This was requested after interviews were conducted and objective sustainability outcomes were collected. A general definition was provided of sustainability to be “the continuation of PCIT” and initiatives were asked to rate how well they felt that PCIT was currently sustaining.

25. Reviewer #2 The description of the qualitative analysis in the Methods seems fine. The authors report the range for the kappas. (Page 9, line 4) 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.

Response: The methods section (Page 9) has been revised to clarify the process of coding to reliability, consistent with the approach used by Bakeman and Gottman (1997).

26. Reviewer #2 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.

Response: We agree with the reviewer’s observation that these ratings are not behaviorally anchored and therefore are not objective. We have added text to clearly outline this as a limitation of the current study (Page 21). We have also included a citation directing readers to the literature on rater bias (Page 19). We have changed our working to refer to these data as “quantitative” versus “objective” (e.g., Page 7). No changes were made to the PSAT, as it is a measure which has demonstrated preliminary reliability (Luke et al., 2014). We modeled the barriers and strategies items from the PSAT format. We did pilot these items with four individuals knowledgeable to PCIT but naïve to the details of ongoing program and defined/clarified any unknown term or details. This was included for 7 of 19 items. Then, for those completing the survey, we offered clarification or support with completion of any items. No participants request additional clarification or support.

27. Reviewer #2 The Results section (Page 9, line 10) 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?

Response: Thank you for this suggestion. To assist with ease on the results section we have moved interviewer data to a subsection following the primary initiative data (Page 17). We agree that analyzing agreement statistically would be a more rigorous test, unfortunately, given the sample size and limited power, we were unable to examine this specifically in the current study. We have attempted to address this by including a section in the results and discussion as well as the limitations of this approach (Page 19). While we appreciate the difficulty in drawing conclusions with discrepant reports, we feel that the inclusion of interviewer ratings is a strength given that most often authors rely on self-report data when examining sustainability outcomes.

28. Reviewer #2 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.

Response: We agree with the organizational comments and have made substantial revisions to the results section (Page 10). We have also added text to the discussion section to suggest potential conclusions related to these differences (Page 19).

29. Reviewer #2 Please possible typo on Page 15 in the paragraph under subheader, "Balancing supply": I believe the first ratings are interviewer ratings.

Response: Thank you for catching this typo. “Initiative” (Page 15) was changed to interviewer ratings.

30. Reviewer #2 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.

Response: Related to the multiple regression analyses to predict sustainability outcomes, we have included text to address the small sample size and statistical power in the limitations section (Page 21). We have also adjusted the text to provide the percentage of potential participants who participated in the current study (Page 7 & 21).
31. Reviewer #2 In the Discussion (Page 17, line 60), the authors cite a sustainment rate of 89% from the Stirman review. I have not located that statistic in the Stirman review.

Response: Due to substantial revisions to the discussion, this text is no longer included.

32. Reviewer #2 The Study Limitations section (Page 20, line 45) needs to be more comprehensive.

Response: As noted in previous comments, we have expanded the discussion of limitations of the current study (Page 21).

33. Reviewer #3: The language "negative child outcomes" seems too specific in this sentence since the paper's patient/intervention focus hasn't been introduced yet. Please rephrase and broaden this sentence.

Response: The introduction text has been revised and this text was removed in this process.

34. Reviewer #3: Please include in the participant description, the number of people involved in initiatives who were contacted but declined participation.

Response: We have added details that no initiatives declined participation (Page 7), while 14 individuals involved in these initiatives were recommended to participate but did not participate. Reasons individuals did not participate included job changes, feeling that others would be better able to report details or that they would not have additional information to contribute, retirement, reported schedule conflicts.

35. Reviewer #3: Please briefly describe how the respondents were selected for the sustainment survey?

Response: We clarified the selection process for individuals who completed the sustainment, barriers, and strategies survey and the PSAT (Page 8).

36. Reviewer #3: Given the wide range of time since initial implementation across these initiatives, interpretation of some data may be more relevant when considering time since implementation (e.g., Barriers such as openness). Was this included as a variable in the regression analyses?

Response: The reviewer question highlights the importance of using time as a variable for analysis. Please see response 21 above, which addresses this point.

37. Reviewer #3: Were survey respondents themselves the identified PCIT champions in some cases? If so, this should be noted and implications discussed.

Response: Yes, PCIT champions were frequently recommended to join the group consultation call. We have added some text to discuss this further (Page 21).
38. Reviewer #3: Appeal was identified as a barrier by the majority of initiatives, however, there are no examples given of low appeal or areas of concern. Please include such examples.

   o Response: The reviewer highlights a finding which we agree warrants further elaboration. We have added additional text to describe examples of barriers related to appeal (Page 13-14).

39. Reviewer #3: The Integration into practice section describes efforts to extend PCIT to new settings and populations. This extension of programming to new populations and settings is novel and encouraging and should be highlighted in the discussion.

   o Response: Thank you. We have added more on this point to the discussion (Page 18).

40. Reviewer #3: Please include in the discussion of Interviewer/Initiative differences in ratings potential biases as well.

   o Response: Please see earlier responses (e.g., 26 & 28); we have added a discussion about these differences (Page 19-20).