Author’s response to reviews

Title: Testing a theory of strategic implementation leadership, implementation climate, and clinicians’ use of evidence-based practice: a five-year panel analysis

Authors:

Nathaniel Williams (natewilliams@boisestate.edu)

Courtney Wolk (cbenja@upenn.edu)

Emily Becker-Haimes (embecker@upenn.edu)

Rinad Beidas (rbeidas@upenn.edu)

Version: 1 Date: 19 Dec 2019

Author’s response to reviews:

Response to Reviews

IMPS-D-19-00515

Testing a theory of strategic implementation leadership, implementation climate, and clinicians’ use of evidence-based practice: a five-year panel analysis

We thank the expert reviewers for their enthusiasm regarding the contribution of our study and their constructive suggestions for improving the manuscript. Below we present the reviewers’ comments and the revisions we made in response.

Reviewer #1:

Comment 1. Abstract: in Conclusions (increase leadership behaviors). Can you clarify here… do you mean an increase in the frequency of leadership behaviors, or some other dimension?

Response: We have revised the Abstract (p. 3) to clarify that we are referring to the frequency of leadership behaviors.

Comment 2. Introduction: I appreciated how the team highlighted the importance of implementation climate in non-routinized settings on page 6.

Response: We thank the reviewer for this positive feedback.
Comment 3. Introduction: Page 7 - the distinction between implementation and transformational leadership is interesting. Can you say more about this? Is there overlap between these two types of leadership?

Response: We have added information on pp. 7-8 describing implementation and transformational leadership and how they work together to influence implementation.

Comment 4. Introduction: The intro might benefit from a little more development of H3 (that changes in implementation climate are not associated with increase of psychodynamic practices). Why is this important for building this theory, etc.

Response: We have added information on pp. 8-9 to further develop H3 and explain why the inclusion of this hypothesis is important for theory testing.

Comment 5. Methods/Results: Could you describe the premise of difference-in-difference approaches a bit more?

Response: We have added information on p. 10 and p. 18 to clarify the premise of difference-in-differences designs and our analytic approach to estimating these effects. In addition, in order to more closely align our study with the classic difference-in-differences paradigm, we have conducted new analyses which we report in the Results section on pp. 21-23 and which we describe on p. 18-19. These analyses incorporate a traditional generalized difference-in-differences approach in which there are two groups at each time point—one that is exposed to the ‘treatment’ (i.e., high implementation leadership/climate) and one that is not exposed (i.e., low implementation leadership/climate). These analyses more concretely demonstrate the effect of change from low to high leadership/climate and significantly strengthen the study findings.

Comment 6. Methods/Results: One of the study's strengths is the longitudinal nature (3 time points), but this raises some questions about whether there is existing evidence about whether/how implementation leadership and climate change over time, how long it takes to change, and what we might expect to observe over a 5 year period? The Table 1 descriptives and Figure 1 suggest that there are actually quite minimal observed changes over time...

Response: We have added information on pp. 7 and 8 to clarify that, to the best of our knowledge, data from previous studies are lacking on how implementation leadership and EBP implementation climate change over time. We have also conducted new analyses to confirm that many organizations in our sample exhibited significant within-organization change in these variables from wave to wave and that these changes were medium to large in magnitude based on Cohen’s d effect size; we describe these new analyses on p. 20-21. We also added information to Table 1 showing the percentage of organizations that exhibited a medium sized change (Cohen’s d \( \geq .5 \)) from wave 1 to wave 2 and from wave 2 to wave 3.
Comment 7. Methods/Results: One of the many things I appreciate about this larger project is the community-wide focus on implementation (over 100 organizations). However, this study only focuses on the 29 largest organizations and I think there needs to be greater detail, explanation of this decision, and perhaps a thorough acknowledgement of some of its limitations. For instance, what did you base your assessment of organizational size on (e.g. annual revenue, # of staff)? Were these private organizations that received public dollars? If so, what percent were private for-profit vs. private non-profit? Why were only 30% of the total organizational population selected? More importantly, what potentially do we lose by excluding smaller organizations? (especially considering how small orgs play a vital role in urban org ecology often as culturally specific, or niche service providers).

Response: We provide additional information on p. 12 to reflect the information in this response. We used a purposive sampling approach for organizational recruitment. Philadelphia has a single payer system (Community Behavioral Health; CBH) for public behavioral health services, thus we obtained a list from the payer of all organizations that had submitted a claim in 2011-2012. There were over 100 organizations delivering outpatient services to youth. Our intention was to use purposive sampling to generate a representative sample of the organizations that served the largest number of youth in the system. We selected the first 29 organizations as our population of interest because together they serve approximately 80% of youth receiving publically funded behavioral health care. The majority of the remaining organizations in the system were very small and did not employ many clinicians and/or see many youth. The organizations that we recruited were geographically spread across Philadelphia county and ranged in size with regard to number of youth served. We have added these details to p 12. We also added to the limitations section on p 28 that these results are likely most reflective of larger organizations rather than single-clinician providers of therapy services as well as the important point that smaller organizations may play a vital role in culturally specific or niche service providers.

Comment 8. Methods/Results: Is there value in describing the % of workers who were the same across waves? A little more detail about recruitment procedures would be helpful.

Response: We have added information on p. 13 indicating the number of clinicians who were present at 1 wave, 2 waves, and 3 waves in the sample. Additionally, we have added more details about recruitment on p. 13. Specifically, with the permission of organizational leaders, researchers scheduled group meetings with all clinicians working within the organizations that delivered youth outpatient services, during which the research team presented the study, obtained written informed consent, and collected measures onsite. The only inclusion criterion was that clinicians deliver behavioral health services to youth (clients under age 18) via the outpatient program. We did not exclude any clinicians meeting this criterion and included clinicians-in-training (e.g., interns).

Comment 9. Methods/Results: Appreciate the thoughtful description of data aggregation on page 15.

Response: We thank the reviewer for this positive feedback.
Comment 10. Methods/Results: Significant control variables might be noted in results section (e.g. clinician experience)

Response: We have added discussion of the significant control variables in the Results section on p. 22.

Comment 11. Discussion: Would like to see significant expansion of the discussion - considering the rigor and the findings here, there should be substantial meat to dig into. For instance what do we learn about change over time and the lifecycle of implementation? Do these small changes mean that leadership and climate are not as malleable as we'd like? Could it be possible that they fluctuate much more frequently than captured in this study? I'm curious what your findings mean for advancing theory in this area. (e.g. I find myself mulling over Aarons theory and Birken middle managers theory which both posit similar mechanisms re: leadership roles/behaviors and climate - can we reconcile them in any way? What else do we need to do to firm theory up here to move our science forward?). What do your findings mean for implementation strategies/practice? What's the significance of H3 in all of this?

Response: We have expanded the discussion section to further explore the implications of our results. On pp. 25-26 we now discuss the implications of our findings in the context of other theories such as Birken et al. and we call for future research in this area. We also discuss the significant within-organization changes that occurred in implementation leadership and EBP implementation climate during the study period and explore the implications of these changes for both practice and research. On p. 24 we revised our discussion of hypothesis 3 to clarify that it helps demonstrate discriminant validity by showing that the relationships we observed were specific to the target behaviors and not generalized to non-targeted behaviors.

Reviewer #2:

Comment 1. I find that the concepts of implementation leadership and implementation climate are well defined, as is the first level leaders' role in shaping EBP implementation climate. In this context, it seems, first level leaders perform a similar function to clinical champions, whose effect has been shown to variable, but with the additional capacity of authority to reward the use of EBP.

Response: We thank the reviewer for this positive feedback and we appreciate this parallel between the functions of clinical champions and first-level leaders. We have added a mention of how this theory overlaps with the concept of clinical champions on p. 25.

Comment 2. Whilst outside of my field of expertise, the quasi-experimental longitudinal differences-in-differences study design and statistical methods appear rigorous and well-justified in the absence of a randomised controlled design and make good use of data collected through the real-world launch of EPIC to improve EBP implementation in a Medicaid-funded
behavioural health provider network. Sampling, data collection methods and measures are adequately detailed.

Response: We thank the reviewer for this positive feedback.

Comment 3. Clinicians average years of experience was noted to be the only time-varying workforce characteristic associated with change in clinicians EBP use. What was the direction of this association i.e. did clinician experience increase or decrease change in EBP use?

Response: We have added information on p. 17 indicating that this was a positive relationship; organizations with more experienced workforces reported greater use of EBP.

Comment 4. A measure of clinician attitudes towards EBP implementation, alongside their perceptions of their organisations EBP climate, may have been a useful addition. Top down support for implementation from the executive or first level leaders may not always be matched with bottom up support for implementation on the ground. Positive clinician attitudes towards EBP implementation may be an underlying factor independent of implementation leadership.

Response: We have added information on p. 17 indicating that in our sample, change in clinician attitudes towards EBP was not significantly related to change in clinicians’ EBP use; consequently, we did not include it in our analyses. We also note that as a robustness check, we re-ran our analyses controlling for clinician attitudes and the results were identical in terms of statistical significance and the magnitude of effects (also, attitudes were not a significant predictor in any model).

Comment 5. As the authors have noted, I find the major limitation of this study to be the use of self-reported use of EBP. Other studies have shown that self-reported practice may not reflect actual practice. I agree with the conclusion that observational metrics of fidelity would increase rigour and potentially an independent audit of a random sub-sample may be a useful addition.

Response: We have added information and a citation on p. 27 noting that self-reported practice may not reflect actual practice and we have strengthened our call for future studies that incorporate observational measures of fidelity.

Comment 6. A useful next step would be to conduct a similar study in a different healthcare system to test generalizability of the findings and of the theoretical model.

Response: We have added information on p. 27 indicating to the importance of similar studies in other healthcare systems to test the generalizability of the results and of the theory.