Author’s response to reviews

Title: Leading for the long haul: a mixed-method evaluation of the sustainment leadership scale (SLS)

Authors:

Mark Ehrhart (mehrhart@mail.sdsu.edu)
Elisa Torres (eltorres@ucsd.edu)
Amy Green (a4green@ucsd.edu)
Elise Trott (ETrott@pire.org)
Cathleen Willging (CWillging@pire.org)
Joanna Moullin (jmoullin@ucsd.edu)
Gregory Aarons (gaarons@ucsd.edu)

Version: 1 Date: 06 Aug 2017

Author’s response to reviews:

Dear Dr. Squires and reviewers:

We are pleased to submit this revised version of our Research manuscript entitled “Leading for the long haul: a mixed-method evaluation of the sustainment leadership scale (SLS)” for consideration for publication in Implementation Science. We appreciate the positive reviewer comments as well as the issues to be addressed. Below, we have addressed all reviewer comments and believe that this process has resulted in a stronger manuscript.

Response to Reviewers:

Reviewer #1: In this study, the authors propose the sustainment leadership scale (SLS) as an extension of the implementation leadership scale (ILS) and assess its reliability and structural validity. Using mixed methods, the authors also corroborate the four dimensions of the SLS and identify a possible fifth dimension: available leadership. The authors find that the SLS has promising psychometric properties and seems to capture aspects of sustainment leadership that are salient to front line providers. The use of CFA and the mixing of methods are key strengths of the study. The authors make an important contribution by highlighting the importance of leadership in the sustainment phase, offering a promising measure for capturing sustainment leadership, and surfacing a previously hidden aspect of sustainment leadership.
My concerns with the study have to do with the meaning of sustainment leadership. It is not clear from the manuscript what the authors mean by "leadership" or how "leadership" differs from management or supervision.

Response: We have added a definition of leadership, as well as some explanation for our use of the term “leadership” versus management or supervision (see page 4). In short, there is little consensus on this point (Yukl, 2013), and thus we use the term “leadership” to be consistent with related literature in this area, including research on focused or strategic leadership such as “service leadership” and “safety leadership” in the industrial/organizational psychology and management literatures as well as the general use of the term “leadership” in frameworks from the implementation science literature.

Likewise, it is not clear what organizational level of leadership the authors have in mind with the SLS. You do not find out until page 9 that the referent for the SLS is supervisors' leadership for sustainment.

Response: We have attempted to be more explicit about the level of leadership that is our focus in this particular study. Specifically, we follow Aarons et al. (2014) in focusing on first-level leaders (i.e., those who supervise direct service providers) in our quantitative study (see page 6). However, in our qualitative study, we did leave the instructions open for comments on leaders across organizational levels (page 13). Nevertheless, the majority of the qualitative data was focused on first-level leaders (page 16), and so we comment on what few insights were gained on the role of upper-level leaders in sustainment and raise that as a direction for future research (pages 21-22).

In addition, there is a brief note that leadership for sustainment is a lot like leadership for implementation, but this idea is not really developed; it is offered as a plausible basis for retaining the four-dimensional structure of the ILS for the measure of SLS but it is not really examined critically.

Response: In our revision, we sought to re-frame this issue as one that is ambiguous, with arguments to be made for why leadership for implementation should be similar to or different from leadership for sustainment. Thus, our design was intended to address both possibilities. We also highlight that leadership questions focusing on implementation are not necessarily appropriate when agencies are in the sustainment phase, providing another justification for the SLS. See pages 6-8 for these additions.

Finally, it is not clear how leadership for sustainment "works" (that is, which outcomes it affects or how it affects them). Some of these issues are discussed in the Discussion section, but they should be addressed earlier since the frame/situate the psychometric assessment and exploration process.

Response: We have added additional information about the foundation of our approach in the literature on organizational climate and culture, and draw from theory and empirical findings in that area to clarify how leadership for sustainment is expected to have its impact through the climates leaders form in their units (see page 5).
Knowing, for example, that supervisory leadership is the focus, at least in this study, of the construct of sustainment leadership would have helped me accept/make sense of the four dimensions the authors identified. These are not the first four dimensions of leadership that come to mind, for example, when I think of sustainment leadership at the top management level.

Response: As noted above, we have edited the manuscript to be more explicit about our focus on first-level leaders in the quantitative data and our attempt to expand across levels of leadership in the qualitative data.

[On a related note, the referent for sustainment leadership in the qualitative component is not specific to supervisors. Study participants’ comments about the sustainment leadership of agency directors and other higher-ups are reported as corroborating evidence for the SLS dimensions. This analytic strategy implies that the concept of sustainment leadership is homologous at various organizational levels. This might be true, or it might not. In either case, it would be nice to see the authors address this question conceptually earlier in the manuscript.]

Response: The purpose of this manuscript was to take initial steps to validate the scale for studying sustainment leadership of first-level leaders. A secondary goal was to examine whether there may be perceptions of differences in sustainment leadership across levels (page 7). As noted previously, almost all of the qualitative inquiry was focused on first-level leaders (page 16), and so we comment on what few insights were gained regarding the roles of upper-level leaders in sustainment and raise that as a direction for future research (pages 21-22).

A related concern has to do with the measurement of sustainment leadership. As I understand it, the authors wish to characterize the behavior of supervisors that contribute to sustained use of an EBP. The authors are not interested individual perceptions (or individual differences in perceptions) of the supervisor's behavior per se; they are interested the supervisor's behavior itself. However, the authors derive data about supervisors' leadership behavior from the multiple employees that report to a supervisor. Like climate, culture, readiness, and other constructs, what we have here, it seems, is a composition variable. In other words, sustainment leadership is a construct that emerges from the responses of individuals within groups. This looks like a referent-shift variable since all members of the team are asked to describe the leadership behavior of the same referent (i.e., their supervisor). If this is correct, then the authors should not be addressing the nesting of providers within team by removing the effect of clustering; instead, they should be measuring the sustainment leadership construct as a group-level construct.

Response: In the leadership literature, leadership is commonly studied at both levels (individual subordinate perceptions and aggregate ratings of the leader behavior) depending on the specific research question and issue being addressed. In the current case, we would not limit the use of the measure to studying aggregated ratings, as there are multiple theories from the leadership literature that suggest that the leader’s behavior can vary within a work group (e.g., leader-member exchange theory [Graen & Uhl-Bien, 1995], and situational leadership theory [Blanchard, Zigarmi, & Nelson, 1993]). Although there are theoretical reasons to support the use of the scale at either the individual or unit level, practically speaking we did not have a large enough sample size at the unit level in this study (i.e., 32 teams) to justify conducting a multilevel confirmatory factor analysis. We have added aggregation statistics to show initial
support for using the measure at the unit level (see page 15), and we have also added material to the Discussion section noting our inability to run multilevel CFA as a limitation and suggesting this issue as a direction for future research (see page 22).


Suggested Major Revisions:

Add a Conceptual Framework to the manuscript in which you (a) define the concept of leadership, (b) note that leadership is a multi-level construct, (c) delineate the organizational level of leadership of interest in this study, (d) define the dimensions of leadership captured in the SLS, (e) ground/justify these dimensions in the broader literature on leadership, (f) sketch out a nomological network that identifies the antecedents and consequences of sustainment leadership, even if only briefly.

Response: We have attempted to address all of these issues in our revision of the manuscript as follows:

(a) We have added a definition of leadership to the Introduction of the manuscript (page 4).

(b) We have highlighted the critical role of leaders across levels in our introduction (e.g., page 5) and generally been more precise about the level of leadership we are referring to when applicable (e.g., page 7).

(c) We have added explicit discussion of the level of leadership that was the focus of past literature on the ILS and the levels that are the focus of this research (page 7).

(d) We have defined these dimensions in the Introduction (page 6); we use the same definitions as proposed by Aarons et al. for the Implementation Leadership Scale (ILS).

(e) We have added material on the development of the ILS and its foundation in the literature on the influence of leaders on organizational climate and culture (page 5).

(f) We have described a general model of how leaders influence the climate of their units, which then influences the motivation and behavior of workers, and ultimately the units strategic outcomes (page 5).

If the construct refers to attributes or behaviors of the leader, rather than employees' perceptions of the leader, consider conducting a multi-level confirmatory factor analysis so that the level of the measurement model matches the level of the construct (or what Klein and Kozlowski, 2000,

Response: As noted previously, there is support from the literature for studying leadership at both of these levels of analysis, but we do not have the sample size at the unit level to conduct a multilevel CFA. We have added this as a limitation and direction for future research (page 22).

Suggested Minor Revisions:

Describe the amount of missing data and indicate whether any pattern of missing data existed.

Indicate whether any multivariate outliers were identified and addressed. CFA is sensitive to multivariate outliers. Please check and report.

Response: There were no missing data within our quantitative sample, thus no pattern of missing data existed. We conducted the Mahalanobis distance test to examine whether there were cases of multivariate outliers. We identified a total of 10 cases that resulted in a p<.001, indicating they were an outlier. In a follow-up sensitivity analysis, we removed the 10 identified outliers and reran the 2nd order CFA. This model resulted in comparable model fit statistics ($\chi^2(50)=110.532$, p<0.001; CFI=.97, RMSEA=.091; SRMR=.033). In the statistical analysis section on page 11, we removed the sentence that specified how missing data were imputed, replacing it with a description of our missing data and our analysis of multivariate outliers.

Revise the abstract for clarity. Some of the sentences are awkwardly constructed (e.g., first sentence the Background and the Methods).

Page 5, line 29: what does "both" refer to here?

Response: We have removed the word “both” to clarify the sentence.

Page 10, line 29: replace "less" with "fewer."

Response: We have made this change.

Page 12, line 39: "strong" is probably too strong a word to describe the fit, given the fit statistics presented.

Response: We have tempered our classification of the model fit statistics, referring to them as “good” fit rather than “strong” fit.

Page 13, line 30: how was "salience" operationally defined?
Response: We operationalized salience as the proportion of codes within each dimension. We have added information on this operationalization to the text.

Page 15, line 5: remove extra period.

Response: We have made this change.

Page 16, line 7: replace "describe" with "described."

Response: We have made this change.

Reviewer #2:

I had the opportunity to review the manuscript entitled, "Leading for the long haul: a mixed method evaluation of the sustainment leadership scale." This study adapted the Implementation Leadership Scale (ILS) to evaluate its ability to characterize key leadership domains of relevance in the sustainment period. The a priori hypothesized four factor structure was supported, indicating structural validity was established, and subscale (as well as total scale) internal consistencies were very strong, indicating reliability of responses was achieved in this sample. The expansion component of their mixed methods study revealed a possible fifth factor reflecting leadership availability that may be particularly important for sustaining the use of evidence based interventions. This study has numerous strengths including the sample size, setting (diversity of contexts, providers, teams), the mixed method analytic approach, and the focus on measure development in the understudied phase of sustainment. The manuscript also has several weaknesses that undermine its contribution to the literature. Weaknesses include the lack of theoretical justification for the measure in its current form (i.e. identical to ILS, but for changes to referent and phase), the lack of a criterion confirming the relevance of these subscales for sustainment, and the approach to qualitative coding. These issues and minor additional weaknesses are elaborated on below.

Major Revisions

1. Most importantly, there is insufficient theoretical justification for the dimensions of leadership and regarding why a unique sustainment phase leadership measure is needed. Without this rationale, it is hard to appreciate the need for this measure, especially because it is identical to the ILS (save for replacement of the referent and phase). Here are some questions and issues that feed into this comment:

Why are the EPIS and CFIR the only frameworks referenced? Consider highlighting the role of leadership across the 60+ D&I frameworks or provide a rationale for why these two are the only two presented in the introduction. Moreover, it does not appear that either of these frameworks informs the development of the leadership domains, antecedents, or related criterion.
Response: These frameworks were the most influential in the development of the ILS, but they were only mentioned as two examples of highly cited/utilized frameworks that highlight leadership and other organizational characteristics. We have edited the opening paragraphs of the manuscript to clarify that reference to these frameworks are meant as examples.

Related to this first point, no nomological network is constructed, nor is domain delineation provided to ensure that the constructs or components of the SLS are sufficiently different from one another and from similar constructs. This is especially relevant as the authors grapple with potential overlap between the "available" construct that emerged from the qualitative analysis and the perseverant and supportive components. Moreover, this process would help the reader appreciate the difference between sustainment leadership and general leadership.

Response: We have made two major changes to address this issue. First, we have provided more detail on the development of the ILS (see page 6), which forms the foundation for the current study. Specifically, we describe the literature review and subject matter expert input that informed the development of the initial item pool, and the multiple rounds of exploratory factor analysis, followed by confirmatory factor analysis, that supported the dimensional structure of the ILS. Second, we have added more detail on the literature on organizational climate, culture, and strategic leadership that is the basis for the expected outcomes of sustainment leadership (see page 5).

The author's conceptualization of "leader" is not well specified. When summarizing the literature in the second paragraph, it is unclear if the data reflects leadership behaviors that could be espoused by anyone on a provider team or if there is a status, role, or hierarchical position that makes these behaviors especially impactful/important? This information (even if conceptual or theoretical would inform with whom one should use the measure in the future).

Response: In this revision, we have made several changes to address this issue. We have added a definition of leadership to the second paragraph of the introduction (page 4). We have clarified that we follow Aarons et al. (2014) in focusing on first-level leaders (i.e., those who supervise direct service providers) in our quantitative study (see page 6). We also note that in our qualitative study, we did leave the instructions open for comments on leaders across organizational levels, although the majority focused on first-level leaders. Thus, we now emphasize the role of upper-level leaders in sustainment as a direction for future research (pages 21-22).

Sustainment is not carefully defined. The second paragraph on page 5 attempts to justify the need for this measure beyond the ILS by discussing both phases (implementation and sustainment), but the authors make a more compelling case for the idea that sustainment is very much contingent on implementation (e.g., "sustainment is, by definition, dependent on implementation").

Response: As noted in our response to Reviewer 1, in this revision we have sought to re-frame this issue as one that has some ambiguity, with arguments made for why leadership for implementation should be similar to or different from leadership for sustainment. Thus, our design was intended to address both possibilities (see pages 6-8).
Although the authors end the second paragraph on page 5 by stating, "the two processes [implementation and sustainment] have unique aspects that may require distinct leadership skills" there is no mention of what these might be. Moreover, their study is not designed to reveal these unique aspects, expect in the qualitative method, which arguably is not set up to yield sustainment specific leadership skills (see point #3 below).

Response: As noted above, we have sought to re-frame this issue as one that is somewhat ambiguous, with arguments to be made for why leadership for implementation may be similar to or different from leadership for sustainment (pages 6-8). We have also added discussion of this issue to the Discussion section on page 23. With regard to the qualitative design, see our response to point #3 below.

Do the instructions for the measure make clear who participants should consider in their responses? Some staff consider "leaders" to be middle or upper managers, or some settings actually have positions called "team leaders" and because the authors had a specific role in mind, it is important to ensure that participants shared their view. Moreover, the authors note in the discussion that leaders at different levels within an organization may espouse different leadership characteristics and engage in different activities to support sustainment. This might be best considered in the background when building rationale for the measure as opposed to included in the discussion. And, although the authors suggest that perhaps the SLS would demonstrate a different factor structure across different leader levels, it may also be the case that different characteristics (and therefore different items) are needed to reflect optimal sustainment leadership, which should be included in their discussion.

Response: Prior to sending out the web surveys, we obtained team structure information for each participating team. Therefore, for each service provider, we were able to include their direct supervisor’s name within the question. This ensured that the staff were providing responses referencing their immediate supervisor. We agree that depending on leadership level, different leadership behavior may be warranted. We have added text on pages 21-22 within the discussion section noting that future research should examine whether additional items are necessary when examining sustainment leadership at the upper-levels of leadership (i.e., executive level).

Although this would be a significant revision, this reviewer wonders if the ILS and SLS were administered in this same study and, if so, could the authors explore the predictive validity of the ILS. Demonstrating the predictive validity and providing evidence of the unique variance accounted for by the SLS would provide important support for this new measure. Making the case for the need for a separate measure is critical given the challenges faced in science and practice with respect to limited opportunity to collect measures from providers.

Response: Unfortunately, the ILS was not administered in this study. Moreover, this comment raises the question of whether a side-by-side comparison of the ILS and SLS would even be appropriate. One of the original motivations for this work was that the wording of measures focusing on implementation would not necessarily be appropriate once implementation is over. In other words, implementation leadership and sustainment leadership are not expected to occur simultaneously, but to correspond the respective implementation phase in which a particular unit
is engaged. Thus, the potential need for the SLS is defined by the lack of any measure that specifically asks about leadership during the sustainment phase (page 6).

2. The authors don't consider alternative factor solutions, likely because their a priori hypothesized model fit the data well, but it seems prudent to consider a single leadership factor.

Response: You are correct in that we chose not to consider alternative factor solutions due to there being an a priori hypothesized model. However, we did run an additional one factor model, removing the four a priori dimensions. This model fit considerably worse ($\chi^2(50)=389.746$, $p<0.001$; CFI=.74, RMSEA=.199; SRMR=.059), thus providing additional support for our decision to examine the hypothesized model.

3. The qualitative questions do not clearly address sustainment. Perhaps this is not an issue because all agencies would be considered in a sustainment phase, but it is concerning that the word "sustainment" does not show up in the focus group questions that were analyzed. The manuscript indicates that these "semi-structured questions addressed positive and negative influences on SafeCare implementation and sustainment." How did the coders distinguish between responses regarding the impact of leadership on implementation versus sustainment? Because of this, it is suggested that the "available" leadership component be referred to as preliminary, especially in light of the fact that no criterion is provided.

Response: The term “sustainment” was not used in the instructions because we wanted to use terminology that was accessible and understandable to participants, and we were not sure of their familiarity with the concept of EBI sustainment. Indeed in implementation science there are multiple terms such as sustainment, sustainability, and maintenance, that all refer to a similar issue. The opening instructions of the qualitative data collection stated that the purposes were to understand “what’s working with SafeCare, the challenges involved in delivering this program, and what it will take to keep this service available in the... system in future years.” Thus, the focus was on how things were currently going with the delivery of the EBI and its ongoing delivery in the future. The combination of the fact that all of the teams participating were in the sustainment phase and were at least 1.5 years (and up to ten years) post-implementation, and that the wording focused on the present and future delivery of the EBI, gave us increased confidence that the answers provided were addressing leadership during the sustainment phase and not during the implementation phase (see pages 12-13). That being said, we concur that without additional evidence, the Available Leadership dimension should be considered preliminary, and we have added wording to the Discussion section to make this clear (p. 21).

4. There is no discussion of inter-rater reliability for the qualitative coding, or how the transcripts were distributed across coders, or what sharing “their work with one another for review” means.

Response: We have attempted to clarify these issues on pages 13-14 of the manuscript. We used a consensus-based approach to determine the final set of codes. As such, inter-rater reliability
was not a primary criterion; instead, we wanted to allow for each coder’s unique insights (Pope & Mays, 2006) that formed the basis for the group discussion and final decisions on the coding.

5. There seem to be additional conceptual differences between the quantitative measure of leadership and the results from the qualitative focus groups. For instance, it seems like the survey subscale of "knowledge" could be enhanced by incorporating knowledge of fit of the EBI with one's workflow or caseload as this came up in the qualitative data, but was not acknowledged as being distinct from the quantitative items. Similarly, it seems that some unique elements emerged in the qualitative data that are subsumed within "supportive" leadership but are quite distinct from the items that comprise that subscale. Specifically, at least one finding in this section seems to suggest that supporting adjustments to workflow/tasks is important for leaders (line 13-15 on page 15). Finally, it seems like there is a unique element emerging in the perseverant leadership section of the qualitative data wherein adapting SafeCare to fit complex patients is useful, but this is not captured in the quantitative measure. It would be helpful to hear how the authors made decisions about how these potentially unique dimensions of each leadership domain were best conceptualized as fitting within the existing domain.

Response: This is an interesting issue, and speaks to the original development and intent of the ILS scale. In line with the additional information on the ILS development that has been added to the introduction section, the scale was developed based on a larger pool of 29 items that were derived from a literature review on leader’s culture/climate embedding mechanisms and strategically-focused leadership, and subject matter expert input. Exploratory factor analyses suggested that a four-factor solution best captured the dimensional structure of the data, and the final set of items was identified based on the strength of the factor loadings and the conceptual alignment with the overall “theme” of the dimension, with the added goal of keeping the scales as short as possible to enhance the likelihood of their use in applied settings. Thus, the goal of the measure development was not to include items that captured all possible manifestations of that particular dimension of leadership, but to include the items that best captured the general conceptual meaning of the dimension. With that in mind, our process for matching the qualitative data with the quantitative data was not an item-by-item comparison, as the items were not intended to capture all manifestations, but instead to consider whether the qualitative data fell within the general definitions of the dimensions or captured something different outside of the dimensional domains. We have added clarification to our description of the coding process to ensure readers’ understanding of these issues (see page 14).

Minor Revisions

6. The authors are encouraged to discuss their intended use of the SLS. Given how the items are worded, it seems like it is primarily an evaluation tool to assess the degree to which leaders espouse leadership characteristics or engage in activities during the sustainment phase to support EBI use. There is much to be said about planning for sustainment and it could be useful for the authors to discuss how their measure might inform sustainment in earlier phases. For instance, given that the ILS is essentially identical to the SLS, the authors might be able to use (either of) the tool(s) to predict sustainment. Or, perhaps building capacity could focus on growing these
characteristics and activities in leadership during the implementation phase. Helping readers to think through when, where, and with whom the tool is useful would be welcomed.

Response: We agree that more discussion of possible uses of the scale was needed, especially for practitioners. As mentioned previously, one motivation for creating the scale was so that items would be appropriate for use depending on the organization’s current phase in the implementation process. Thus, we would argue that the primary consideration for deciding which scale should be used would be current phase of the organizational unit. We would expect that units with leaders scoring high on the ILS would be in a better position for sustainment, especially since they would be expected to have had more effective implementation, which is the foundation for sustainment. However, with changing priorities of organizations, it is certainly possible that strong implementation leaders could not continue to be strong sustainment leaders. It would be of interest for future research to address what factors contribute to when strong implementation leaders continue as strong sustainment leaders and when they do not. To address these issues, we have added a practical implications paragraph at the end of the Discussion section (page 23).

7. On page 7, line 8-12, is the bulk of the sentence referring to coaches or supervisors?

Response: The demographic information provided pertains to the service providers who participated in the web-based survey. No supervisors or coaches participated in the web-based survey.

8. The two sentences beginning on line 12 and going through line 24 are very confusing. Please revise as it seems like both are missing words.

Response: Thank you for your suggestion, we have modified those sentences to ensure they flowed better.

9. It would be helpful to report the distribution of years implementing SafeCare across the participating agencies and not just providing the range of 1.5 to 10.

Response: The distribution is as follows: Two systems were implementing for 1.5 years, one system for 2.5 years, three systems for 3.5 years, and one system for 10 years. This information has been added on page 9.

10. Please provide the correlation among the factors.

Response: We have calculated the interfactor correlations for the four subscales and provide these in the manuscript on page 15.
Discretionary Revisions

On page 6, line 29, consider replacing "address" with "contribute to."

Response: We have made this change.

Consider replacing "analyses" with "setting" on page 7, line 34.

Response: We have made this change.