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: 2 Date: 30 Nov 2017

Author’s response to reviews:

We thank the reviewers for their helpful comments. Below we list each comment and our response.

Reviewer #1

The authors have been responsive to the questions and concerns I raised in my earlier review. They have offered a definition of leadership, situated their study within the broader leadership literature, and clarified the level of leadership to which the measure applies. The authors also signaled how sustainment leadership "works" (i.e., what sorts of outcomes sustainment leadership would or could influence). Given the close connection between theory and measurement, the manuscript would make a more substantial contribution to the field if the construct of sustainment leadership were situated within a well-argued nomological network. However, the revised manuscript moves further in this direction than the original manuscript.

I appreciate the authors' efforts to address my concerns about the level of the construct. The authors state that they do not have enough teams (N=32) to support a multi-level confirmatory factor analysis. I'm not expert enough in multi-level confirmatory factor analysis to confirm that 32 teams is insufficient for such an analysis. However, I do know that there are no hard and fast rules; we have rules of thumb that are helpful, but statistical power in multilevel confirmatory factor analysis, as in many other statistical techniques, depends on several factors,
not just sample size. In circumstances such as these, I recommend consulting with a psychometrician with expertise in multilevel confirmatory to determine whether this approach to assessing structural validity is feasible with these data.

Response: We did indeed consult with our study statistician, an expert in multi-level analyses, for the last revision. He confirmed our own assessment that there were not enough level-2 units to conduct a multilevel CFA.

Regardless of whether multilevel confirmatory factor analysis is feasible with these data, the broader concern about the level of the construct remains. I strongly recommend that the authors state whether they view sustainment leadership as a property of the leader, a property (perception) of the followers, or a relational property between leaders and followers. This is crucial for defining what the construct of sustainment leadership is and, obviously, how the construct of sustainment leadership is measured. If we conceive of sustainment leadership as a property of the leader, then each leader has a sustainment leadership score. If we have multiple followers rating the leadership qualities of the leader, we would expect some level of agreement among the followers about the qualities of leadership the leader they follow demonstrates. Let's take Knowledgeable Leadership as an example. If half of the people following a leaders say she's knowledgeable and half say she is not knowledgeable, what does that tell us about the leader? We can't say she's a Knowledgeable Leader, nor can we say she's not a Knowledgeable Leader. How then can we characterize her Knowledgeable Leadership? As an exercise in arithmetic, we could compute a mean by averaging the followers' perceptions, but that mean is not a valid depiction of the leaders' Knowledgeable Leadership, nor is it even an accurate reflection (representation) of the views of the followers. We can't say she's a kinda-sorta, middle-of-the-scale Knowledgeable Leader, nor can we say that her followers think she is a kinda-sorta, middle-of-the scale Knowledgeable Leader. This is not an academic point. The authors note that the ICC(1) for the SLS is .05, meaning only 5% the variation in sustainment leadership scores occurs between leaders. The rest of the variation in sustainment leadership scores occurs among followers rating the same leaders. It would appear that followers following the same leaders view those leaders differently. While the awg(j) statistic for the sample as a whole is a respectable .73, the low ICC(1) value suggests that there are many teams where inter-rater agreement is not so high. The authors could check this by calculating awg(j) for each team.

An alternative to viewing sustainment leadership as a property of the leader is to view sustainment leadership as a property (perception) of the follower. In other words, sustainment leadership is in the eye of the beholder. We distinguish organizational climate (a property of the workplace) from psychological climate (a property of employees). A similar distinction could be made for leadership. This would put sustainment leadership in a different nomological network, one organized around the causes and consequences of individual perceptions of leadership. Since this is the level of measurement the authors employ in this study, I recommend the authors state explicitly that they view sustainment leadership as a perception of followers and note that followers following the same leader can have different perceptions of the leader's sustainment leadership. The authors have already noted in the revised manuscript that the factor structure they obtained holds at the individual (follower) level and that the factor structure (number of factors and factor loadings) could differ at the leader (supervisor) level. This is worth reinforcing
so that readers do not leap to the conclusion that they can compute SLS scale scores for leaders using the structural validity information obtained at the individual level of measurement/analysis.

As an aside, leader-member exchange theory focuses on the relationship between leaders and followers and notes that leaders can have different relationships with followers. Here the level of the construct is the dyad, not the leader. So, the level of the construct (theory) and the level of measurement align. Even for theories of situational leadership—which argue that leaders have behavioral flexibility to lead in different ways in different situations—we would still expect followers in any given situation to rate the leader in similar ways (e.g., democratic leadership), even if followers in another situation rate the leader in a different way (e.g., autocratic leadership).

Response: This is an interesting issue, and although we agree with some of these points, we disagree with others. We have two main responses:

1. Although we agree that the ICC(1) value is not very high, we do not entirely agree with the assessment given of the agreement statistics. First, although an ICC(1) value is not large, we would not dismiss it too quickly. We refer to this quote from LeBreton and Senter (2008):

   “Specifically, a value of .01 might be considered a ‘small’ effect, a value of .10 might be considered a ‘medium’ effect, and a value of .25 might be considered a ‘large’ effect (see Murphy & Myors, 1998, p. 47). For example, an ICC(1) = .05 represents a small to medium effect, suggesting that group membership (e.g., employing organization) influenced judges’ ratings (e.g., employees responses to a climate survey). Thus, values as small as .05 may provide prima facie evidence of a group effect” (p. 838).

   Thus, the .05 value could be considered a small to medium effect and adequate support for the presence of a group effect. With regard to the interpretation of the ICC(1) value in combination with the awg(j) values, we again refer to LeBreton and Senter (in which they consider ICC(1) to be in the category of IRR+IRA and IRA to be agreement measures like awg(j)):

   “LeBreton et al. (2003) demonstrated how it is possible to have high levels of IRA yet low levels of IRR and IRR + IRA. They showed that when between-target variance becomes substantially restricted, correlation-based estimates of IRR and IRR + IRA are attenuated. In such instances, researchers relying solely on ICCs to justify aggregation could make very erroneous decisions” (p. 840).

   In other words, ICC(1) values can be low due to low within-group agreement OR low due to between-group variability (or some combination of the two). In this case, the agreement levels meet standards for “strong agreement” in LeBreton and Senter (p. 836), and although there were some groups for which agreement levels were lower than others (note that we now include the range of values in the text), the factor that is most likely the primary cause for the lower ICC(1) values is low between-group variability. Thus, we would conclude that there is adequate evidence based on the combination of these statistics for aggregating the SLS to the unit level, and we would suspect that in other samples with more variation between groups, that the ICC(1) values would be higher. At the same time, we agree that it is important to clarify that we have
not tested the factor structure of the construct at the aggregate level; because it is possible that the factor structure could vary at the group level, future research should build on our initial findings to address this issue. We have added a brief summary of some of these points in the paper to clarify this issue.

2. Regarding the level of the construct, our measure of sustainment leadership was designed in line with most measures of leadership in the organizational behavior literature. The focus is on the attributes and behaviors of the leader as perceived by subordinates, much in the same way that transformational leadership or other behaviorally-based leadership constructs are studied. Although these measures address attributes and behaviors of leaders, they are used to study leadership at multiple levels of analysis. For example, transformational leadership is considered a “meso” theory that cuts across levels of analysis (Wang et al., 2011). Even as much of the transformational leadership literature either addresses individual perceptions of leaders or mean levels of perceptions at the group level, there is also growing research acknowledging that there can be variability within groups, such that there is agreement in some groups and differentiation in other groups, and there are multiple factors that act as predictors of that variability in addition to differing consequences of that variability (e.g., Feinberg et al., 2005; Nielsen & Daniels, 2012). We believe our approach to sustainment leadership is consistent with much of the leadership literature, and that it would be unnecessarily limiting to specify a level of analysis from the outset when sustainment leadership could be studied at multiple levels using the measure we developed (in the same way that transformational leadership and related constructs are studied). In other words, we believe that there are likely to be some situations in which subordinates will agree about the leader’s sustainment leadership, and others in which the subordinates will vary in their perceptions of the leader’s sustainment leadership. Researchers should specify how they are conceptualizing sustainment leadership in their particular study based on their specific research questions, and then their use of this measure should be aligned with that conceptualization. We have added some additional discussion of this issue to the Discussion section.

To summarize:

Please note that leadership can be conceived and measured in different ways (as noted above) and specify upfront the level of the sustainment leadership construct.

Response: We have added information that sustainment leadership can be conceptualized and studied in multiple ways, and that researchers should ensure that the level of theory for their specific study is aligned with the level of analysis.

Please acknowledge the meaning (implications) of the low ICC(1) value, which reinforces the value of measuring sustainment leadership at the individual follower level.

Response: We have added information that the lower ICC(1) value is most likely due to low between-group variability, that the initial results presented here support the aggregation of the SLS measure to the group level, and that future research should examine aggregation statistics of
the measure, particularly in situations in which the levels of sustainment leadership are expected to vary more between groups.

Please add the caution that the structural properties of the SLS could differ at the leader level from those observed at the follower level.

Response: We have added such a caution to the Discussion section.

Reviewer #2

I had the opportunity to review the revised manuscript titled, "Leading for the long haul: a mixed-method evaluation of the sustainment leadership scale (SLS)." The authors did an excellent job revising the manuscript in response to reviewer concerns, notably in clarifying their measure development process, situating their measure in the broader literature, and describing the important need for this measure in addition to the ILS. Simultaneously, although the authors build the case for the distinct need for a measure of sustainment leadership, and yet their approach to developing it (using the theory, research and process of developing the ILS and then changing referents) does not support their claim. Moreover, their mixed methods approach seemed to undermine the utility or meaning of the data collected, and there remain conceptual issues with their construct operationalization and its alignment with the measure items.

Major Revisions

* This reviewer does not feel the response to major issue #5 from the initial review was sufficiently addressed. If the authors did not wish to use the qualitative data to validate or expand on the conceptual clarity and appropriateness of the quantitative measure then it's unclear what purpose the qualitative data served. The integration of mixed methods feels quite selective. That is, it seems the authors have missed an opportunity to refine the SLS based on their good qualitative data, but are choosing instead to go with their literature review on the ILS to confirm its validity, which somewhat contradicts the authors' statements about ILS and SLS needing to be quite distinct.

Response: Taking a step back, it seems that these comments suggest that we should have taken a more traditional mixed methods approach to instrument development, so we want to acknowledge that our approach may seem “unconventional” in that it doesn’t fit one of the primary existing mixed-method designs. However, our goal was to use methods that were appropriate for our specific situation and questions. As mixed-method researchers are increasingly acknowledging, there are situations that may require quantitative and qualitative data to be integrated in ways that do not cleanly fit within standard approaches (e.g., Poth’s recent IIQM webinar entitled “Mixed Methods Research Complexity: Dilemmas and Opportunities”). Thus, we have added the following material to the Method section on pages 14-15:
“We acknowledge that our design deviates from the traditional sequential-exploratory mixed methods approach to instrument development in which the research topic is first explored qualitatively with participants, with qualitative findings then used to develop items for quantitative psychometric testing. Our study, however, employed a different approach as our first goal was to determine whether an existing measure of implementation leadership could be modified to measure sustainment leadership using the same factor structure. Our secondary goal was then to use the qualitative data to further validate the dimension of sustainment leadership with a group of service providers sustaining an EBP and to expand upon whether additional themes related to sustainment leadership emerged.”

We also wanted to respond to some of the specifics in these comments as well. Throughout our manuscript we do not argue that the implementation leadership and sustainment leadership are “quite distinct,” but instead that there has been little attention on sustainment leadership in the literature, and there are reasons to think that the behaviors associated with each have some aspects in common and some aspects that are unique (see pages 6-7 of the manuscript). As such, one of our goals was to examine the extent to which a measure already validated for measurement of “implementation leadership” would also maintain strong psychometric properties and factor structure when used to assess leadership for sustainment. We used the qualitative data to validate the quantitative measure by examining the extent to which the dimensions were represented in those data, and assessed whether the measure needed to be expanded by identifying any additional dimensions that were not captured by the original set of dimensions. Our integration of mixed methods was not “selective” as the approach described was in line with our original goal. We sorted the qualitative data into the themes/dimensions originally identified in the ILS research, and we sought to identify any behaviors that did not fit within that original dimensional structure. Our goal was not to create new items with the qualitative data; we had already shown in the quantitative analyses that the items could be used to capture sustainment leadership. In the original ILS work, a large set of items was created with the feedback of subject matter experts, and through exploratory factor analyses, was narrowed down to a narrow set of items (three per dimension) that would capture the “spirit” of the dimension in an efficient manner. Multiple items were removed that captured unique behaviors, but the goal was not to create an exhaustive list of unique behaviors, as that would result in an unwieldy, impractically long measure. We maintained that approach in this paper; the goal was not to add individual items capturing unique sustainment leadership behaviors, but to first show that the ILS items with the changed referent would “work” for measuring sustainment leadership, and second to ensure that there were not dimensions of sustainment leadership not captured by the original dimensional structure. If the CFA for the original set of sustainment leadership items had not been supported, then it would have been appropriate to start from scratch with item creation. But because those items were supported, creating new items was not necessary. Thus, as we stated in the original response, our focus was at the dimensional level, not the item level. Although the reviewer may disagree with this approach, we believe that our approach was reasonable; due to the overlap in implementation and sustainment, it seemed to make sense to benefit from past work in this area and use that as a foundation, in addition to recognizing the possibility that differences that may exist, and thus using the qualitative data to extend the dimensional structure of the instrument.
Similarly, the authors seem to argue for the distinct predictive utility (or simply composition) of sustainment leadership and yet they were very much reliant upon the ILS to build the sustainment measure with no additional theory, implementation relevant expertise, or empirical work brought to bear in the introduction or discussion.

Response: We disagree with this representation of our work. We believe we were very straightforward about building on the work from the ILS and provided reasonable arguments for why that approach was taken. Our approach is clearly described in the paper on page 7:

“In light of the arguments to be made for both the overlap and distinctiveness of implementation leadership and sustainment leadership, we used a research design that addresses both. Specifically, we first used a quantitative approach to test whether the dimensions of the ILS hold in the context of sustainment, and then used a qualitative design to connect descriptions of leadership behaviors during sustainment to that same dimensional structure. Then, to address the possibility of differences in critical leadership behaviors in the two different phases, we identified any new themes in the qualitative study that did not align with the original proposed SLS dimensions.”

Although the authors have now adopted a published definition of leadership, it is clear that their scale items do not derive from this definition. Accordingly, there remains concern about the construct validity of the resulting measure. The "embedding mechanisms" discussion on page 14 more closely aligns with the items and so it may actually be useful to rethink the specific construct they seek to measure in that it may not broadly be "sustainment leadership." This may seem like a minor point, but our field struggles with conceptual clarity and the authors have great potential for shaping the field in methods of measure development.

Response: Our inclusion of a general definition of leadership was included at the request of the other reviewer; it was included as background for our discussion of leadership. We agree that we did not include a definition of sustainment leadership, and that has now been added to the paper on page 7 (“the attributes and behaviors of leaders that support the effective sustainment of EBI implementation”). This definition is consistent with our discussion of the construct in this paper, with the literature Schein’s embedding mechanisms, and with other research on strategically-focused leadership.

Minor Revisions

* The word "framework" shows up twice in a row in the methods section of the abstract.

Response: Our original version did not show this error, but we have made sure it is correct in this version.

* The word "of" is missing from line 46 on page 13.
Response: Unfortunately, we read through this several times and were not able to locate where the word “of” was missing, but hopefully any error along these lines would be identified during the copyediting process.

References not already included in the manuscript:

