Author's response to reviews

Title: Should researchers use single indicators, best indicators, or multiple indicators in structural equation models?

Authors:

Leslie A. Hayduk (lhayduk@ualberta.ca)
Levente Littvay (levi@littvay.com)

Version: 3 Date: 2 October 2012

Author's response to reviews: see over
Reply to Reviewer 1

Reviewer’s report
Title: Should researchers use single indicators, best indicators, or multiple indicators in structural equation models?
Version: 2 Date: 10 July 2012
Reviewer: Christian R Geiser

Reviewer’s report:
I have read the authors’ responses to my comments and the new version of their manuscript carefully and with great interest.

We thank you for your conscientiousness as a reviewer. We expect there will be many others who will, like you, want to read and re-read this article with “great interest”. We feel this article is worthy of the attention, and will reward the readers’ efforts.

I appreciate the authors’ detailed explanations in their response letter.

Our responses to the prior rounds of review tried to respect and respond to the substance of each and every reviewer comment, and the same applies to our current responses. We hope all six reviews and responses become available when this article is published.

A number of things have now become clearer to me with regard to the authors’ way of thinking. Also, the manuscript has improved, particularly the introduction, which is now more (albeit still not fully) understandable to readers not familiar with the related SEMNET debates.

We agree that there was some improvement in readability and clarity, though the substance of the arguments remains unaltered. We address this reviewer’s comment regarding SEMNET in the context of his additional comments below.

Compared to current practices, the paper offers a radically different perspective on selecting and specifying indicators in structural equation models (SEMs) with latent variables.

We agree that our paper presents a “radically different perspective” that offers a strong alternative to many people’s “current practices”.

It calls for a more rigorous theory-driven approach to setting up SEMs with which I strongly concur.
Our approach is more theory-based, and it encourages easier/clearer connections between researchers’ theories/thinking and their data/analyses.

The paper makes a number of very important points and should be useful to advance the field in several ways.

Thank you for these comments.

One of my remaining concerns is that, given the radical differences between the approach advocated by the authors and the approach currently pursued by the “mainstream” (including both methodologists and substantive researchers), the current write-up is still too much targeted towards the specific audience that follows (or is at least somewhat familiar with) the related SEMNET debates. As an example, the description of the close vs. exact fit issue on p. 3 is cryptical and definitely hard to follow for somebody not involved in the corresponding SEMNET exchanges. Is this not a side issue for the current indicator discussion? Is it at all relevant here? Could it not be dropped?

We have attempted to minimize the references to SEMNET, but these references cannot be “dropped” entirely because the SEMNET archive documents relevant history demonstrating the interconnections between three measurement issues confronting factor analysis, namely:

(1) measurement using multiple indicators should be done with (not “before”) using a full SE model;

(2) all SE models should be carefully tested (including multiple indicator factor models);

and (3 – our article’s demonstration of how current factor analytic practices can be improved upon) measurement of latents in SE models requires more theory than a factor analytic appeal to common-factors, and specifically requires “prior” specification of the few-best indicators.

Our article only addresses the last of these issues, but the interconnection between these issues is important context. We have slightly-revised the segment of our article that refers to SEMNET to further clarify the relevant context.

Also, I found that although the authors explained many things very well in their reply to my initial comments, most of these very useful explanations did not find their way into the revised manuscript.

We reviewed this reviewer’s prior comments and checked that any explanations relevant to a wide audience have been included in the text. Fortunately, we were able to make minor wording changes that simultaneously address this point as well as the next point the reviewer made (below). Several of the reviewer’s prior comments are different renditions of points we address in the article. Adding these into the article would
constitute repetition – which seems counterproductive since our prior revisions systematically eliminated repetitious material to satisfy the editor’s request. This reviewer’s prior comments, and our prior replies, might provide alternative wordings of interest to some readers but we think the BMC archiving of the reviewer comments and our replies should suffice. We considered adding a comment pointing to this reviewer’s comments and our replies, but given that we were “gently chastised” (by both current reviewers) for pointing to specific reviewer comment and our responses, we decided to simply permit the BMC review archive to function as open-context available to the reader. At the editor’s request, we would be happy to include a statement pointing specifically to this reviewer’s comments and our replies.

The authors now mention “misconceptions” briefly at the end of the manuscript (p. 33) and it sounds a little bit like those were just my personal “misconceptions”. But their explanations come at an odd place and are not at all integrated into the general discussion. It is awkward to say, “OK, were done with the main discussion now and here are some of Christian Geiser’s misconceptions.” Moreover, the explanations given on p. 33 are way too sparse for the reader to fully understand the issues. I am fairly certain that the "misconceptions" that I raised in my previous review are not just my personal ones. In fact, many other researchers in the field share these “misconceptions”. So I think that it would be necessary and worthwhile to address these points in the manuscript in the same helpful way in which the authors did this in their response letter. This would help readers to better understand why the authors’ approach is seen as superior over currently established procedures.

We agree, and we have now deleted this section and incorporated all the points at appropriate places within the text.

Regarding the "misconception" of breadth of latent variables, my point was that many researchers at least in psychology would not accept a latent variable labeled, say "depression" that was operationalized with just a single item, say "I feel sad all the time". First, single items usually do contain more random measurement error than do aggregates based on multiple items. Of course, the meaningful use of an aggregate requires the assumption that the items are at least congeneric (and an appropriate weighting scheme if the items are not at least essentially tau-equivalent), but with more than 3 items, this assumption can be tested. Many people would think that if we can find a set of congeneric items, it would generally be better to use the set of items rather than a single item, because the aggregate would be less contaminated by random measurement error.

Unfortunately no single variable (latent or otherwise) has any “breadth”. Variables have magnitude, but no breadth. Our article makes it abundantly clear that the assumption of a latent common-cause can be tested even with two or three indicators – more than three indicators are NOT required for such testing. Notice also that such testing requires that all the items be included in the model, and not just the “aggregate”, so this reviewer
is either implicitly arguing for a two-step process of measurement-before-latent-path-model (namely the first of the 3 problematic factor analytic claims listed above and referenced with published works coordinated with SEMNET) or is arguing that the measurement be assessed along with the latent level of the model (which moves the reviewer toward agreement with the view presented in our article).

Notice that the “aggregate” the reviewer refers to would have to be modeled as a single indicator with a fixed/specified amount of error variance, though this could NOT be done appropriately by the reviewer unless the multiple indicators were assessed in the context of a full latent level model and NOT just a factor-structured model. With a factor-structured model, the available error variance information would concern reliability and NOT validity. To claim validity the reviewer would have to appeal to assessment of the multiple indicators in the context of other theory-relevant latents. If the reviewer wishes to add the requirement that the multiple-indicator assessment be done in the context of other theory-relevant latents, that brings the reviewer one step closer to agreeing with our position.

Second (and probably more importantly), one basic idea of factor analytic measurement models (I know the authors are not a big fan of factor analysis, and I’m not saying that I am one) is that random measurement error AND specific factors can be separated from common factor variance by using MULTIPLE indicators of a more broadly defined construct, say depression. Call it breadth or whatever else, but to me it seems that this is seen as one of the major strengths of the SEM approach and of using MULTIPLE indicators by many people in the field. Again, I’m not saying that I share this view. However, this point is likely of interest to the readers, and a main point of controversy. Unfortunately, it is not sufficiently addressed in the current version of the manuscript.

We have now indicated how “specific factors” can be usefully addressed. The issue is not the breadth of a variable – the real issue is the possibility of multiple causes of a variable, where there are many more options than the multiple-causes being factor-structured. The possibility of multiple causes is implicit in the notion of error variables as net-causal-actions of omitted variables (plural) though there is no reason to presume that “specific factors” are the most common interference because there is no reason to presume that the interfering features are routinely factor structured. Some people might presume that specific causes must be specific-factors, but we view that as a disciplinary bias rather than veridical assessment. We have clarified how specific “factors” can be addressed in the context of fixed measurement error variances.

Additional, more specific comments:
- The authors largely ignored my previous comments on the reliability-validity discussion. Even though they make it clear in their response letter that “the issue is validity, not mere reliability”, this still does NOT become clear in the paper. The authors still use expressions like “true-score-like latent variables”, “true score”,...
and “true-score variance” (e.g., on p. 3, p. 12, and p. 14) and “measurement error” (in various places in the manuscript, e.g., on p. 14 and p. 25), which to many readers will suggest that the issue is reliability as described in classical test theory (CTT), from which these terms are borrowed and which focuses on reliability, not validity.

We have now clarified that our use of such terms is not to invoke reliability or the CTT thinking, but to refer to the world’s true (even if unknown) causal structuring.

As another example, on p. 13, the authors state that “attaining a valid model specification might require specifying y5 as less reliable.” (emphasis in the original) As I had tried to make clear in my first review, this will likely be confusing for readers, because typically, a correction for attenuation is based on CTT-based measures of reliability, and NOT validity estimates.

This unfortunate use of the term “reliable" has been corrected.

So what the authors propose is something completely new and unusual, and hence terms should be used in a very clear and consistent manner. Furthermore, as an issue raised previously, Borsboom and Mellenbergh cautioned against mixing up issues of reliability and validity on both the conceptual and empirical level, and this issue is still not discussed by the authors.

We are in close agreement with Borsboom et al. on multiple issues, and we have now inserted statements clarifying that we are concerned with validity, and not mere reliability. What we propose is not “completely new and unusual” – portions of the basic claims appeared in Hayduk 1987, 1996 (as we indicated in our references), though we understand that this may appear as “completely new” to those who have not encountered the relevant works.

- I insist that the way the authors currently discuss their “real examples” (starting on p. 28) is rather unhelpful for researchers who want to understand how to derive reasonable estimates for the fixed error variances in practical applications. Especially the current description of the loop simplex model (Example 1) is completely useless, as the authors do not explain anything related to the fixed indicator residual variance issue for this example, but instead only refer to a book chapter.

References 37 and 38 are both to published articles and NOT a book chapter. This example was introduced to show how even identical replicate measurements can sometimes require latents that are NOT a common factor, or common-cause. We hear
these statements as also requesting additional assistance with error variance specification, which we address in the context of the reviewer's next comments.

I can understand the authors' statement that every research situation is unique and that general guidelines are therefore hard to give. But the authors must have gotten their ideas for how to fix the residual variances in their examples from somewhere and even describing that would be helpful, because it gives researchers an idea of how experts proceed in these situations, even if this cannot be directly transferred to their own research problem. Otherwise, the "real examples" will just fill space, but will be of little or no utility to interested readers.

We understand the reviewer's request for additional direction, and have now incorporated some suggestions for how to proceed. We did not insert this in the "real examples" section, because it seemed to fit more naturally elsewhere (namely in the section on how to specify measurement error variances).

Level of interest: An article of importance in its field
Quality of written English: Acceptable
Statistical review: No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests:
I declare that I have no competing interests.
Response to Reviewer 2

Reviewer's report

Title: Should researchers use single indicators, best indicators, or multiple indicators in structural equation models?

Version: 2 Date: 17 July 2012

Reviewer: conor dolan

Reviewer's report:

Minor Essential revision required.

Review by CVDolan of “Should researchers use single indicators, best indicators, or multiple indicators in structural equation modeling” by Hayduk & Littvay (revision).

Revision ok. Interesting & thought-provoking paper, welcome discussion in SEM (whether you agree or not with the present thesis).

---

We agree with “interesting & thought-provoking”, and anticipate some controversy – because important changes do not happen without controversy.

---

Here are some residual remarks and a few minor details. I did get the cover letter, but it did not include detailed responses to my previous remarks. (I emailed about this, but did not get a answer). No matter, except I might repeat myself.

---

We provided detailed responses explaining the changes we made in response to your prior comments, but we do not know why you did not receive these.

---

comments & minor details

The authors have decided to address Geiser's comments at the end of the paper. Let us suppose that Geisser offered comment, which was interesting, but wrong. Then the comment should be incorporated – properly embedded – in the text to avoid the apparent confusion. If the comment was simply ill-conceived, farfetched and wrong, then it can be dealt with in the detailed responses in the cover letter (which we did not get). It is most unusual – and not particularly nice, actually – to make a point of Geisser’s “misconceptions” at the end of the paper.

---

We have now deleted the section on misconceptions and incorporated the relevant comments into the body of the article.

---

I may have asked this before: what does this imply for the psychometrician who develops a unidimensional psychometric instrument on the basis of the usual techniques (IRT modeling, assessment of dimensionality, iic’s, etc.)?
Any technique that actually locates a unidimensional causal-latent should be consistent with use of a single indicator of that latent in structural equation modeling. Investigating why multiple IRT items might, or might not, function as multiple indicators in structural equation models is worth investigation but would warrant a whole new article because this raises interesting (and we think unresolved) issues about how or whether IRT latents connect to causal latents.

p. 4 line 6 from bottom to p. 6. Quite wordy. I suppose that this is necessary, given the readership of this journal? Yet notwithstanding all these words in explanation, one aspect of fig 2 is not broached. The residuals are now “fixed”. I understand that this will be explained later on, but would it not be easier on the reader to devote a sentence to this on page 5 (“note that...we return to this later on”). Or this information could be added to the figure 2 caption.

This has been done.

p. 4. line 2 ....Figure 1attended...
p. 6 line 5 from bottom: Tthe...

Both of these typo’s have been corrected.

p.6 line 4 from bottom. “The latent-effect portion of the model contributes importantly to eta3’s identity or meaning, and only causes of y5 other than eta3 constitute error”. Not exactly an easy sentence to read.

We have altered, and hopefully improved, this sentence.

p. 8. line 16. ...this implies. ... this implies:

The parallel sentence constructions here were intended to highlight the parallel logical sequencing of the equations.

Figure 3: Please add to the caption why the regression coefficients connecting the etas equal 1.

This is explained in the text just below equation 7 (which we have revised slightly). The figure caption did not seem like the proper place for an explanation of why the 1.0’s appear.

Also read as rendered, this Figure stipulates that eta3a = error-A. That is the intended reading?
Yes, but remember that Figure 3 is depicting options for a segment of Figure 2, where there will be effects entering and leaving whichever version of eta3 (whether eta3a, eta3b or eta3c) is selected as the appropriate version of eta3. The text makes clear that where other effects enter or exit from eta3 is an important yet open/undetermined issue. Consideration of the open options is crucial in determining which version of eta3 should be incorporated into the model. Latent effects leading to or from all three versions of eta3 (including eta3a) cannot be included in this diagram because there has been no specification of which of the three latents is the one that should receive and/or send latent effects. Figure 3 is therefore unavoidably unspecified in this regard. Fortunately, this diagrammatic sparseness helps highlight the important point regarding error accumulation. The text and both Figure 2 and Figure 4 progressively modify the optional readings of Figure 3.

p. 12. line 11. “Similarly detailed assessments should accompany each fixed measurement error variance in the model (‘)”. One the one hand the authors require many words to explain relatively simple concepts, on the other aspects of their modeling approach are suddenly mentioned without any appreciable introduction. This makes the paper hard to read. As far as I can tell this is the first mention of “fixed measurement error variance” in the text (fixed appears in Fig. 2). At this point, no explanation has been given of why one should fix the error variance. I understand that this will be discussed later on, but the reader may wonder about this when he/she encounters this on page 13.

We have now introduced the notion of fixed error variances earlier in the article (page 5), and this is specifically addressed on pages 8 and 9 where this is connected to the strategy for selecting between the optional meanings for eta3.

Am I correct that ultimately “Hayduk’s procedure” is outlined on page 17 top? So the reader encounters “fixed error variance” in fig 2 and on page 12, and then – confused perhaps – has to read patiently on to page 17 to understand the purpose of these fixed parameters? It is more likely that the reader is confused by this lack of explanation and simple puts the article aside.

We have now introduced the idea of fixed measurement error variances earlier in the article – on pages 5, 8, and 9, as well as 12, 16, and so on.

p. 30 last line: on the basis of what the researcher thinks they know...of what researchers think they know

The English of this sentence has been improved.

**Level of interest:** An article whose findings are important to those with closely related research interests
Quality of written English: Acceptable
Statistical review: No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests: Nothing to declare