

# Statistical and Methodological Myths and Urban Legends

## Where, Pray Tell, Did They Get This Idea?

Robert J. Vandenberg  
*University of Georgia*

Although a kernel of truth or appropriateness underlies each one, numerous methodological and statistical criteria have been stretched beyond that kernel and have evolved into myths and urban legends. They are truly legends in that those applying it were told it was so and have thus accepted it and perpetuated it themselves in interactions with others. This is particularly prevalent in the manuscript development and review processes in which the mythical aspects of those criteria are applied. The end result is a degradation of the research process. The feature topic's purpose is to uncover the kernel of truth underlying the mythical nature of alternative model testing, cutoff values for certain criteria, common method bias, and mediation test. It is developmental in nature in that the desired result is to end the legendary application of those criteria and return to applying them in their intended manner.

**Keywords:** *statistical myths; alternative models; research process*

The use of the term *statistical* in the title is purposeful in that the focus of the current feature topic is on the myths and urban legends as they pertain to the quantitative approach to the research process and not on those myths or legends that may characterize the qualitative approach. The overall process of research regardless of one's primary approach is a difficult one to say the least. As this process pertains to a general quantitative approach, one must first master the conceptual literature of whatever substantive phenomenon she or he has chosen and from that mastery convince others that there exist researchable gaps in that literature. Second, assuming such gaps do exist, it is improbable that all can be addressed simultaneously in one grand study, and thus, the researcher must take a programmatic approach. Third, it is understood that as the researcher undertakes that program, new substantive questions (i.e., gaps) will arise as old ones are addressed. Fourth, addressing each gap requires applying appropriate research methods that validly mimic the conceptual conditions specified by the researcher to represent the variables and their interrelationships. Finally, assuming that the methodology is appropriately operationalizing the conceptual premises characterizing the selected research gap, it is necessary to use a statistical framework that will produce interpretable results while simultaneously accounting for the properties of the data (e.g., scaling, reliability, no independence among the variables, etc.).

Although the above statements imply that there is a clear road map underlying the research process, this impression could not be further from the truth. The steps to the research process as presented in the first paragraph are collectively analogous to a satellite picture of some landmass. Although the details defining that landmass may not be discerned from that partic-

ular picture, we all know that underlying this picture are myriad countless paths, highways, streams, and other features. If we were to show two people from the same geographic location the same satellite picture and ask them to roughly sketch the best way between two points, we would very likely be given two overlapping yet unique sketches. When asked to share their rationale, we would learn that the selected path was based partly on legend (e.g., my “grandpappy” always went this way) and partly on fact (e.g., the interstate in this area is the fastest way around that city). Similarly, two researchers with the same substantive interest would in all likelihood take overlapping yet different methodological and statistical approaches if the same research question were placed in front of them. The research process is so difficult, therefore, because no step as outlined in the first paragraph is characterized by a clear set of guidelines that breaks down the process into a set of easily executable “if-then” decision-making points or criteria. Indeed, it is highly probable that a particular criterion today (if it existed) that is held by researchers as being “the” way to go about a step in the research process will change during the course of one’s career as new knowledge is gained. Hence, the very process of undertaking research is by nature a dynamic one requiring constant vigilance and continuous learning.

The doctoral student training and the journal peer-review processes are the two most predominant mechanisms for traversing and learning the principles of research. During the course of their training, we expose doctoral students to the satellite picture and to numerous guidelines and examples of how researchers (mostly published ones) traversed to various points on that picture. Although some things to learn about the research process become fairly obvious, particularly those to avoid, the impression students most often complete their program with is that there is a lot of “gray” in applying the knowledge dictating the research process. For example, we would no more draw a path on a satellite picture that would ultimately lead one to drive over the lip of the Grand Canyon to certain death as we would teach our students that a measure with a reliability coefficient of .35 is adequate, and thus, they can use that measure in a research context. The grayness comes from the fact that there are different paths around the Grand Canyon, and analogously, different levels of so-called good reliability. The peer-review process is designed to not only reinforce what was learned in graduate school but also to further develop the researcher’s understanding of the research process. That is, the editors and reviewers critique the submissions not only to reinforce good practices but also to point out a weakness in the research process or to direct the author to an alternative aspect of the research process that may be better or stronger in some way than that applied originally by the author. The editorial board members are in a position to do so because they have presumably stood the test of time and established themselves as experts in their own rights both substantively and methodologically.

An unfortunate component of both the doctoral student training and the peer-review process as described above may be summed up in the commonly used phrase “We are only human,” meaning that the grayness leaves a great deal of room for individual interpretation. Doctoral students may be taught or told something to do within the research process as if it were an absolute truth when in reality it is not, and yet, being who they are, they accept that presumed fact as the “truth.” Similarly, authors may accept something from an editor or a reviewer who in turn was told that “this” is the way it must be as well. The unfortunate outcome is that the truism being perpetuated is anything but true. These are aspects of the research process that are, in reality, myths or urban legends. At one point, there may have been a kernel of truth to it, but that kernel has long been forgotten or altered in such a way as to be

lost. Rather, unbeknown to the student, author, reviewer, and editor applying the criterion, it is a criterion of the legendary kind (i.e., “my grandpappy . . .”). There are all kinds of deleterious side effects to this, not the least of which may be the unfair evaluation of a manuscript against criteria that are mythical in nature or the application of the criteria in undertaking some aspect of the research process resulting in a finished study of questionable quality. The overall end result, however, is a degradation of the whole research process.

The purpose of this feature topic is to address the above issue but only as it applies to Steps 4 (applying appropriate design methodology) and 5 (using an appropriate statistical framework) as outlined in the first paragraph above. The general concern is that it is very common to hear of an author whose manuscript was rejected because of an unwarranted application of some methodological or statistical criteria by an editor or a reviewer. Another is to hear an author state that he or she used “such-and-such” criteria, implying that the methodological or statistical quality of the manuscript is now good, but in reality, the application of such criteria is largely myth. The common ones focused on within this feature topic are characterized by the following often-found statements regarding a manuscript: (a) “You didn’t test for any alternative models,” (b) “Your within-group correlation was less than .70,” (c) “Your self-report measures suffer from common method bias,” and (d) “Your test for mediation failed because your X and Y were not significantly correlated.” Although these four are the focus here, there are certainly many others that could be addressed as well: “You have an unmeasured variables issue,” “There is no point in interpreting your main effects when their product is statistically significant,” “You cannot meaningfully interpret the product term because it suffers from multicollinearity,” “Your item-to-subject ratios are too low,” “You can’t generalize these findings to the real world,” “Your fit indices are too low,” and “Your effect sizes are too low.”

The fact remains that there is a kernel of truth to each of the above-listed criteria, but the user (usually unknowingly) is not applying them according to that underlying kernel. That is, although it is extremely difficult to trace events back to when they first occurred, some misconception regarding the criteria began, and more important, became accepted, as fact. It is this misconception and misuse that is referred to here as the myth or urban legend. Given the grayness inherent in the research process, it is easily understood why this may occur. Namely, in their mythical or urban legend status, they help navigate the grayness because they are being used as guidelines that break down the research process into a set of easily executable “if-then” decision-making points or criteria (i.e., if the manuscript fails the criterion, reject it, or alternatively, if I, as an author, apply the criterion, it makes it appropriate). I will briefly use testing alternative models (point a above) as an illustration.

Specifically, it is not unusual for manuscripts using structural equation modeling (SEM) to be rejected, with one of the primary reasons stated by the reviewers and the editor for the rejection being a failure to test for alternative models. Yet on reading the manuscript, it is very apparent that the author has a completely sound conceptual footing for that model. The author is frustrated because to him or her, testing an alternative parameterization among the latent variables is tantamount to engaging in exploratory analyses (a legitimate concern at times; MacCallum, Roznowski, & Necowitz, 1992). Furthermore, when I have had the opportunity to follow up on the reviewers or editor to get clarity on that reason, it is not infrequent for them to state, “Well, that’s how I was taught to evaluate SEM models, and thus, if the author does not provide an alternative, then I heavily advise for rejection.” If I further ask them whether they had an alternative substantive parameterization in mind when they wrote

the comment, the answer is all too often “No.” My reason for asking that question is to determine whether the reviewer possessed some expert understanding of that substantive literature that the author may not have been aware of and that could be reasonably considered in creating a theoretically driven alternative model. In short, I am attempting to discern whether the decision resulting from the use of that criterion was based on its myth or its reality. I undertake a similar inquiry when evaluating my doctoral students’ manuscripts or assisting colleagues by conducting a prereview of their papers. Again, it is not uncommon in those manuscripts to see an alternative model specified, and that alternative is a clear “stretch” relative to the sound conceptual footing underlying the model of interest. When I ask why they included the alternative, it is once more not infrequent for them to state that they know that they will be criticized by the editor and reviewers for not having one, and thus, they are undertaking a “preemptive strike.” As above, my question is stated to discern how much of their decision was driven by the myth versus the reality of the alternative model criterion.

Thus, what is the reality underlying the test of alternative models? Very briefly, it is rooted in the fact that by defining a strong-fitting model via a statistically nonsignificant chi-square value, we are not accepting that model but rather are only failing to reject it. Even under the very best of circumstances in which we obtain a statistically nonsignificant chi-square and the cutoff values for our selected absolute (e.g., standardized root mean square residual), relative (e.g., incremental fit index, Tucker-Lewis index), and adjusted (e.g., parsimonious normed fit index) fit indices have been reached, the truth is that we could take the same set of latent and observed variables and identify a number of equally well-fitting models, each representing a different type of parameterization among the latent variables (Loehlin, 2004; Schumacker & Lomax, 2004). However, although these models are mathematically equivalent, this should not be confused at all with being conceptually equivalent; that is, the model of interest is grounded very strongly in theory, and the others may be discovered only by exploring. Hence, our willingness to conditionally accept (recall that this is not possible statistically) the model is a conceptual decision.

The problem, though, is that very rarely in the organizational sciences do we achieve the “very best of circumstances” as defined above. It is most often the case that the only statistical test of fit, the chi-square test, is statistically significant, whereas the other “nonstatistical” fit indices all indicate strong fit. Technically speaking, the only statistical test, the chi-square, is the one we should attend to, and it is telling us that we should absolutely reject the conceptual model because it is not fitting the data well (Bollen, 1989). Practically speaking, however, it is well documented that the chi-square test is very sensitive to sample size, and thus, very small differences between the observed and reproduced covariance matrices will result in a statistically significant chi-square value when the sample is large (Bollen, 1989; Lance & Vandenberg, 2001). Thus, the common practice among most researchers is to see if the chi-square value is reasonable (yet still, unfortunately, significant) and base the strength of the model on the other indices of fit. The problem with this practice, though, is we are now failing to reject our model on nonstatistical criteria (note this is in part why Hu & Bentler, 1999, called for stricter cutoff values on these criteria). Furthermore, this also means that there could be any number of alternative models that fit mathematically equivalent to, and now also better than, the conceptual model. Hence, once more, our willingness to conditionally accept (again, with the understanding that this is not the same as statistically accepting) the model is a conceptual decision.

In closing this example, the main point from the above material once more is that captured in the reality underlying failing to reject our model. This truly does mean that there is quite possibly a number of mathematically equivalent or even stronger models underlying our data but representing different parameterizations among the latent variables. This is the reason why individuals such as James, Mulaik, and Brett (1982) stressed early on the need to anchor latent variable models in a very strong conceptual foundation because it is the theory that will define an acceptable model from among all possible models underlying a particular database. Thus, part of the alternative-models myth to be avoided by authors is to stop immediately thinking that the fit indices are facilitating your decision to accept your model. Once more, they facilitate the decision only as to whether you can fail to reject the model with the understanding that there are very possibly many equivalent models that could mathematically fit your data just as well as your conceptual model. Therefore, another part of the alternative-models myth to be avoided by authors is to stop denying the existence of those possible equivalent models. Doing so in part means that the conceptual foundation underlying the focal model is so convincingly developed in the first place that any other alternative parameterization is highly improbable. It also means being open-minded and willing to conceptualize and present an alternative model or models to the one of interest when there may be competing views as to how the latent variables are associated with one another (see Vandenberg & Lance, 1992, for an example). This may not necessarily mean that the whole model is reparameterized, but there may be competing views as to how two, for example, of the latent variables in the model are related or perhaps not related to one another that can be supported from the research literature. The philosophy of science literature discusses tests of alternative models under the more general label of *strong inference*. Strong inference refers to undertaking a critical study in which one theoretically driven parameterization among the latent variables is compared to a second or even a third theoretically driven parameterization. It is one of the most powerful tools we possess to pit theories against each other and discard those that do not fit the data well (see Aguinis & Adams, 1998, for an excellent example of this approach). The bottom line is that as authors we are expected to pay diligence to the alternative model criterion both conceptually and statistically, particularly because the “best of circumstances” as defined above is rarely achieved. The latter issues are not new, and systematic approaches for undertaking that diligence were published more than 20 years prior by Bentler and Bonett (1980) and James et al. (1982) and are reemphasized in contemporary books on SEM (e.g., Maruyama, 1998, p. 248). I would encourage authors to review those sources.

Similarly, I would encourage editors and reviewers to attend equally to the above sources as well. As noted previously, there is the reality underlying failing to reject the model of conceptual interest. In this case, though, recall that the mathematical equivalence of alternative models does not mean conceptual equivalence. The alternative-model myth to be avoided by editors and reviewers is to stop immediately using the alternative model criterion as a primary reason for rejecting a manuscript when the authors simply do not present an alternative. For one, that is simply not true in the vast majority of cases because the reason why fit indices such as the normed fit index, Tucker-Lewis index, and the relative noncentrality index (among others) are classified as relative fit indices (Maruyama, 1998; Tanaka, 1993) is that they are computed relative to an alternative model, the null model (a model that is emphasized both by Bentler & Bonett, 1980, and James et al., 1982). Most authors report at least one of those indices or some other relative fit index, and consequently, some part of the alternative model criterion has been met. The strongest reason for avoiding the use of the alternative



model criterion as a primary reason for rejecting a manuscript, though, is in the case in which the authors have done an excellent job providing a compelling conceptual rationale for the parameterization among the latent variables but as an editor and/or reviewer, you simply do not like the manuscript. In those circumstances, articulate the reasons why the manuscript does not fit in that journal but do not hide behind the excuse that an alternative model was not tested. Doing so is a disservice to the authors who may avoid immediately turning around an otherwise excellent manuscript to a more fitting journal. In those cases, though, in which there is truly a conceptually driven alternative model that may be addressed by the authors but they are unfamiliar with some part of the literature, then this should by all means be called to their attention in the form of a criticism. However, if this is the only major weakness to an otherwise strong manuscript, then I would encourage developmental feedback to the authors pointing out what the alternative may be and some of the sources they could use in developing that alternative. The primary point here as it was above is that diligence must be observed as well by editors and reviewers when applying the alternative model criterion. The myth to avoid is to use it “as a hammer to nail a manuscript” (a comment that I have had three individual reviewers and one editor state to me in person on different occasions). The diligence expected here is no less conceptually driven in that if this criterion is evoked to evaluate a manuscript, then inform the authors as to what the conceptual alternative may be and what sources they may consult to develop that alternative. That is, the fact that there are mathematically equivalent or better models should not be confused with the existence of conceptually equivalent or better models. Otherwise, authors may go down some unnecessary exploratory path if they are simply told that “a major weakness of this study is a failure to examine alternative models.”

As discerned from just this one example alone, the underlying message from it is that no matter how much we as researchers desire having very clear “if-then” criteria to apply within the research process, that desire is itself largely a myth or urban legend. That is, grayness is an inevitable characteristic of the total research process, and adopting criteria without fully understanding their roots results as mentioned previously in a degradation of the research process. This message is repeated throughout the three feature articles. In the first article, Lance, Butts, and Michels (“The Sources of Four Commonly Reported Cutoff Criteria: What Did They Really Say?”), for example, separate the myth from the reality of four commonly cited cutoff values (e.g., the  $r_{wg}$  should be .70 or greater). Their point, as was the point above, is that the stringent acceptance of those values is largely an urban legend. As such, the decisions resulting from those legends may not be wholly appropriate and may have resulted in less than accurate interpretations of research findings or in the handling of manuscripts during the review process. The second feature article by Spector (“Method Variance in Organizational Research: Truth or Urban Legend?”) similarly examines another commonly evoked concern among authors, reviewers, and editors: the presence of method bias in a study. As with the first article, Spector visits both the kernel of truth and the mythical nature of method bias. He notes that the issues underlying method bias are in reality quite complex but that the term *common method variance* (CMV) has served to mask that complexity. That is, when the CMV label is used, it is often done so with only one form of bias in mind and, as such, pushes us into using it as if it were an “if-then” criterion of the type described above. Finally, the third feature article of this series by James, Mulaik, and Brett (“A Tale of Two Methods”) examines the realities and myths that have emerged in the context of tests for mediation. They note that over the years, authors have typically adopted one of two methods (either the so-called Baron

and Kenny approach or an SEM approach) in examining mediation hypotheses. The myth or “tale” in their use, though, is that researchers commonly treat these as if they were the same, leading to similar conclusions. The reality, though, as covered by James et al. is that although overlapping in some respects, the two methods are quite dissimilar. The point once again from their article as it is in the other articles is that mediation tests are, in reality, complex and not a set of “if-then” criteria.

In closing, there are many more methodological and statistical urban legends and myths than are presented in this series. Even if coverage of all potential myths had been granted at this juncture, the conclusion of each article would have been the same; that is, seeking to simplify the research process by creating a checklist of “if-then” criteria is itself a myth. I would like to personally thank Larry Williams, the outgoing editor of this journal, for encouraging me to create this feature topic. More important, I thank him for carrying through with his vision in creating this journal. I would also like to thank the authors of the three feature articles for putting their thoughts down on paper. They did so enthusiastically when approached with this idea. Finally, articles are always shaped by the thoughts of the reviewers as well. Therefore, I would like to thank all the reviewers for this feature topic: Gilad Chen, James Conway, David Kenny, James LeBreton, Michael Lindell, and Roger Millsap.

## References

- Aguinis, H., & Adams, S. K. R. (1998). Social-role versus structural models of gender and influence use in organizations: A strong inference approach. *Group & Organization Management*, 23, 414-446.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.
- Hu, L. T., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models and data*. Beverly Hills, CA: Sage.
- Lance, C. E., & Vandenberg, R. J. (2001). Confirmatory factor analysis. In F. Drasgow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (pp. 221-256). San Francisco: Jossey-Bass.
- Loehlin, J. C. (2004). *Latent variable models: An introduction to factor, path, and structural equation models* (4th ed.). Mahwah, NJ: Lawrence Erlbaum.
- MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, 111, 490-504.
- Maruyama, G. M. (1998). *Basics of structural equation modeling*. Thousand Oaks, CA: Sage.
- Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10-39). Newbury Park, CA: Sage.
- Vandenberg, R. J., & Lance, C. E. (1992). Examining the causal order of job satisfaction and organizational commitment. *Journal of Management*, 18, 153-167.

**Robert J. Vandenberg** is a professor of management in the Terry College of Business at the University of Georgia. He is the past division chair of the Research Methods Division of the Academy of Management. He has served on the editorial boards of the *Journal of Applied Psychology*, *Organizational Behavior and Human Decision Processes*, *Organizational Research Methods*, and the *Journal of Management*—journals in which many of his publications have appeared as well—and is currently an associate editor of *Organizational Research*.

*Methods.* His substantive research interests include high-involvement work processes, organizational commitment, and employee work adjustment processes. His methodological research interests include latent growth modeling, multilevel structural equation modeling, and measurement equivalence and invariance. His 2000 publication with Charles Lance on measurement invariance received the 2005 Robert B. MacDonald Award for Best Publication in Methods sponsored by the Research Methods Division at the Academy of Management.