**Statistical and Methodological Myths and Urban Legends in Strategic Management Research: the Case of Moderation Analysis**

This paper examines whether methodological precedence in applying moderation analysis to strategic management research relies on myths and urban legends, and if doing so affected empirical conclusions, implications for theory development, and practical recommendations. An in-depth analysis of 69 studies published in the *Strategic Management Journal* between 2000 and 2014 using moderation analysis finds that strategic management scholars typically rely on statistical myths and urban legends when applying moderation analysis including (1) interpreting main effects separately from their significant interaction with other variables, (2) failing to report reliability values of interaction terms and (3) relying on hierarchical approaches that can lead to interpretation errors. Further examples illustrate how these practices could lead researchers to draw incomplete and possibly inaccurate conclusions. Overall, problematic precedents have become the gold standards for testing and interpreting moderation models. Best practice recommendations for redirecting future research to more solid methodological grounding are provided.

Keywords: Methods, Moderation, Moderators, Myths, Regression

When justifying their methodological decisions, researchers tend to use the common practices in a field as precedent and support (Boyd, Gove, & Hitt, 2005b; Mizruchi & Fein, 1999). Some of these practices become widely accepted without question and develop into gold standards that authors and reviewers rely on in making decisions. However, as methodological practices diffuse, they can evolve into what have become known as “statistical and methodological myths and urban legends” (SMULs) (Vandenberg, 2006). SMULs involve statistical processes and decisions that, although possibly originating in sound science, become distorted and exaggerated as the approaches diffused throughout the field (Spector, 2006: 222) such that the intended meaning was lost along the way (Vandenberg, 2006). Lance has described such myths as “those rules of thumb, maxims, truisms, and guidelines for research conduct…received doctrines…that are otherwise passed from generation to generation. They establish normative research conduct…inform us about how we should go about our research…often based, in part, on sound rationale and justification but also, in part, unfounded lore” (Lance, 2011: 280-281)[[1]](#footnote-1). These statistical rules of thumb, conventional standards, and received methodological doctrines may not be valid (Lance, 2011) and their application can lead to inappropriate findings and inferences (Lance & Vandenberg, 2009; Vandenberg, 2006).

We consider the presence and implications of SMULs in strategic management research which is a relatively young academic field (Boyd, Bergh, Ireland, & Ketchen, 2013). Developed since the late 1960s, it is a multidisciplinary area of research and has borrowed heavily from economics, sociology, psychology, political science, evolutionary ecology, philosophy, and other disciplines (Durand, Grant, & Madsen, 2017). Additionally, the domain of strategic management has been affected by, and in turn affected, other fields of business management such as innovation, marketing, human resource management, business ethics, entrepreneurship and organizational behavior (Durand, et al., 2017; Nag, Hambrick, & Chen, 2007). SMULs, although first discussed among organizational behavior scholars (i.e., Lance, Butts, & Michels, 2006; Vandenberg, 2006), might have a significant impact on strategic management research as well given the multidisciplinary nature of strategic management research. Yet the presence and potential implications of SMULs are unknown: What is the prevalence of SMULs in strategic management research? Have they impacted the interpretation of empirical findings and the implications they have for theory development and practical recommendations? Until we assess their existence, we do not know whether the strategic management research process and the field’s underlying knowledge base are vulnerable to the impact of SMULs.

Although there are many possible SMULs (e.g., Lance, 2011; Lance & Vandenberg, 2009; Lance & Vandenberg, 2015; Vandenberg, 2006) that influence strategic management research, the present study focuses on SMULs of testing moderation since moderation analysis is so common in strategic management, appearing as one of the field’s earliest methodological treatises (e.g., Venkatraman, 1989), and now spanning most major theoretical perspectives in the field (Boyd, Takacs Haynes, Hitt, Bergh, & Ketchen, 2012). There has been little research examining the extent to which moderation analyses have been applied appropriately within strategic management. In fact, to date, only Boyd and colleagues (2012) have carried out a systematic review of how strategic management researchers test moderation. They document the prevalence of moderation analysis, note the proportions of statistically significant results, the percentage of studies that recognize statistical power, and the proportion that report reliability for interaction terms, apply mean centering, and use graphs for visual representations of findings. While insightful at documenting many methodological decisions on moderation, Boyd and colleagues (2012) do not examine whether SMULs exist within moderation analysis and how they impacted findings and inference.

The present study examines the prevalence of SMULs of testing moderation within strategic management research and their implications. Following other assessments of strategic management research practices (e.g., Bergh and Fairbank, 2002; Hamilton and Nickerson, 2003), we conduct an in-depth analysis of articles appearing in the field’s flagship empirical journal, the *Strategic Management Journal* (*SMJ*). Findings from an assessment of 69 *SMJ* studies reveal that strategic management researchers have generally adopted myths and urban legends when justifying key methodological decisions for testing moderation: 1) hypothesized conclusions are often based on simple main effect coefficients when the model indicates conditional relationships; 2) researchers almost universally ignore the reliability of interaction terms (which frequently fall below accepted levels), and 3) hierarchical regression models have frequently led to misinterpretations of variable coefficients. We demonstrate how such practices have led to incomplete and possibly incorrect inferences and conclusions that influence theory development and recommendations. Importantly, we provide best practice recommendations for redirecting future work in strategic management research and social science research at large. They serve as alternative approaches to help create new precedents going forward and for editors and reviewers to judge the publication worthiness of journal submissions.

**MODERATION ANALYSIS AND SMULS**

Moderation occurs when the relationship between an independent and dependent variable depends or is conditional upon a third variable, usually known as a moderator or moderating variable and denoted as a Z (Aiken & West, 1991; Carte & Russell, 2003; Dawson, 2014). In the absence of a moderator, *X* is expected to predict *Y*. However, the moderator term *Z* denotes that the strength of the *XY* relationship will vary depending on the level of *Z*. Indeed, *Z* could theoretically amplify, dampen, or even reverse the effect of *X* on *Y*.

Testing for moderation is typically conducted using multiple regression by introducing the product of the predictor variable *X* and the moderator variable *Z* (*X* \* *Z* = *XZ*) into the regression equation as in the following equation (for more details, please see J.R. Edwards, 2009):

*Y* = a + b1*X* + b2*Z* + b3*XZ* + e. (1)

When *X* and *Z*, the main effects, are controlled, the coefficient b3 represents the change in the effect of *X* on *Y* for a unit change in *Z* (Aiken & West, 1991; Cohen, 1978). A value of b1 (the coefficient representing the impact of *X* on *Y*) at a given level of *Z* is referred to as a simple effect. Boyd and colleagues (2012: 288) note that this approach is now the norm in strategic management research, as scholars “frequently test moderation using interactions of predictor variables (e.g., cross-product multiplicative terms), which are added to a regression model”. Moderator Z can be a categorical variable, in which case moderation effects are tested either by interaction terms as described, or by conducting subgroups analyses (Boyd, et al., 2012; Dawson, 2014). In the former case, dummy variables are created for the categories of the moderator variable and interaction terms are created between X and those dummy variables (Dawson, 2014). In the latter case, *t*-tests of the correlation coefficients are often used to show differences between two groups and chi-square tests are used to show differences among several subgroups (Boyd, et al., 2012).

**Conventional Practices of Moderation Analysis: The Emergence of Statistical Myths**

Edwards (2009) identifies several common methodological practices that can “lead researchers to make unwise choices, waste time and effort, and draw conclusions that are misleading or incorrect” (p. 143).[[2]](#footnote-2) Although Edwards (2009) identifies seven myths relating to moderation analyses, not all of them are equally concerning. Some of the practices rooted in myth, such as the mean centering of variables involved in interaction terms (Edwards’ Myth 1), are relatively benign. Adherence to the practice creates a little extra work for researchers, but does not adversely affect their empirical outcomes (e.g., Echambadi & Hess, 2007). Others, however, are very problematic. For the current purposes of our paper, only those myths that could potentially cause researchers to draw conclusions that are misleading or incorrect, and therefore materially impact knowledge development in the field, are of most interest. These myths include Myths 2, 3, and 4 as identified by Edwards (2009) (see Table 1 for the full set of Myths, recommendations and references related to those Myths). Below we term these three myths as Myth A, B, and C respectively.

**-------------------------------**

Insert Table 1 about here

**-------------------------------**

**Myth A: Interpreting coefficients of the first-order terms**. Some researchers have declared the coefficients representing the direct effects of *X* and *Z* “arbitrary nonsense” (Cohen, 1978: 861) in light of a statistically significant interaction. The basis for this position is that “the coefficients on *X* and *Z* when *XZ* is included in the equation has been a source of confusion…[which] emanates from the fact that, with *XZ* in the equation, the coefficients on *X* and *Z* are scale dependent, such that adding or subtracting a constant to *X* changes the coefficient on *Z*, and vice versa” (Edwards, 2009: 146). In short, the myth is believing one cannot interpret the X and Z coefficients in the presence of a significant interaction without understanding two important caveats. One, the coefficient for *X* represents the simple effect of *X* on *Y* *but only when Z is equal to zero*. If zero is a meaningful value, or *X* and *Z* have been mean-centered to make it so, then the simple coefficients do contain useful information. The second related caveat is that researchers must be aware that the simple coefficients on *X* and *Z* reflect the effect of those variables *only* when the other is equal to zero. At any other value of the other variable, the effect of *X* or *Z* will be different due to the moderation effect. Therefore, general theoretical interpretations about the presumed main/simple effect of *X* or *Z* on *Y* must also include the interaction.

**Myth B: Measurement error of interaction term can be ignored**. Edwards (2009) reports a practice among researchers whereby it is assumed that if *X* and *Z* exhibit adequate reliabilities[[3]](#footnote-3), the reliability of *XZ* is likewise adequate. The end result is that the reliability of *XZ* is not computed, and therefore, its measurement error on estimating the coefficients is not considered. However, “this issue does not simply ‘go away’ when the separate reliabilities for *X* and *Z* are themselves quite strong…” (Edwards, 2009: 158). For example, if *X* and *Z* are uncorrelated, and each has a reliability of .7, then the reliability of *XZ* is only .49. If the correlation between *X* and *Z* increases to .25, the reliability of *XZ* is .52 based on following equation (Bohrnstedt & Marwell, 1978; J.R. Edwards, 2009):

Correlation of XZ-squared + (reliability of X times reliability of Z)

Reliability of XZ =  (2)

Correlation of XZ-squared + 1

Therefore, even when the reliabilities of *X* and *Z* meet some desired level, the reliability of *XZ* can fall well below that level. Results become vulnerable to problems associated with measurement error, which can adversely influence the interpretation of interaction terms.

This myth apparently only applies to measurements that allow reliability tests to be conducted such as multi-item Likert scales widely adopted in organizational behavior research (Edwards, 2001) and non-economics social science. Strategic management scholars often rely on single-indicator measurements which do not allow the assessment of reliability (Boyd, et al., 2013; Boyd, et al., 2005b). Nonetheless, there are some strategic management studies which indeed develop and use multiple measurements which allow reliability tests (Boyd, Gove, & Hitt, 2005a) and thus are vulnerable to Myth B.

**Myth C: Moderation analysis should be tested using a hierarchical approach**. The hierarchical approach is used when direct or main effects and controls are included in a base model, and then the interaction terms for the moderators are added in subsequent models (Aiken & West, 1991). Statistically significant coefficients on the multiplicative interaction terms and changes in the *F*-ratios associated with adding the interaction terms in the subsequent “full” models supposedly provide evidence of moderation effects.

However, this procedure is subject to two drawbacks. “First, when a moderating effect is captured by a single interaction or product term, such as *XZ…* hierarchical analysis is unnecessary because the *F*-ratio…will give the same result as the *t*-test of the coefficient for *XZ…* A second drawback… is that it can generate interpretations of the coefficients for *X* and *Z* that are misleading” (Edwards, 2009: 150-151). More specifically, researchers often interpret the first-order main effect coefficients in the base model before adding the interaction term *XZ*. They then interpret *only* the interaction term in the second step, ignoring the first-order variables. This process invites errors, as the base model interpretations are “unconditional, such that the effect of *X* on *Y* is viewed as a constant across levels of *Z*, and likewise, the effect of *Z* on *Y* is viewed as a constant across levels of *X*. However, if the coefficient of *XZ* is significant in the second step, then the effects of *X* and *Z* are both conditional, such that the effect of each variable depends on the level of the other variable” (Edwards, 2009: 151).

Strategic management scholars that apply the hierarchical approach invite unconditional interpretation of direct effect coefficients in the base model. Making the problem worse, those first-order coefficients in the base model are likely to suffer from omitted variable bias since a theoretically relevant factor, the interaction term representing the moderating effect, is not included. Concluding that a hypothesis is either supported or unsupported based on a potentially biased coefficient in an incomplete empirical model can lead to serious inferential problems. Therefore, “[n]o conclusions can be drawn about main effects (based on main effect coefficients in the base model) in the presence of [significant] moderating effects (in the full model)” (Carte & Russell, 2003: 495).

**SMULS OF MODERATION ANALYSIS IN STRATEGIC MANAGEMENT RESEARCH**

To assess whether the above myths might be present within strategic management research, we identified a meaningful set of strategic management moderation studies by examining all articles published in the *Strategic Management Journal* *(SMJ)* over two adjacent time periods, 2000-2009 and 2010-2014. Nag and colleagues argue that the publication of an article in *SMJ* is seen by strategic management researchers as “providing *prima facie* evidence that it was an SM [strategic management] article” (2007:938). We do not intend to criticize any specific journal, but rather chose this journal simply because the research practices adopted by this leading outlet are likely to be adopted by others. There is also precedent using *SMJ* as a single source for evaluating methodological practices in strategic management studies (e.g., Bergh & Fairbank, 2002; Bergh, Sharp, Aguinis, & Li, 2017; Shook, Ketchen Jr, Cycyota, & Crockett, 2003).

We selected a random sample of 25% of *SMJ* articles that included empirical tests of moderation using interaction terms published between 2000 and 2009 and between 2010 and 2014. This sample provides an opportunity to assess the practice of moderation both before and after the Edwards’ (2009) chapter on SMULs pertaining to moderation was published, and is large enough to permit generalization to the population of published or forthcoming articles in which moderation may be tested. A manual examination of each *SMJ* article during these two periods resulted in the identification of 242 articles which reported tests of moderation between 2000 and 2009 and 50 articles between 2010 and 2014. This led to a random selection of 60 articles between 2000 and 2009 and 13 articles between 2010 and 2014. Four of the 60 articles between 2000 and 2009 did not report sufficient data for evaluation. The final sample included 56 articles published between 2000 and 2009 and 13 between 2010 and 2014, which made it 69 *SMJ* articles in total in the final sample. Each article was assessed with respect to the three myths identified above. The coding process and reliability tests are reported in the appendix.

We next present the findings from our analysis. Further we also provide one example study from our sample to illustrate whether the three myths have a meaningful effect on findings and theory development. We want to emphasize that this illustration is not attempting to highlight inappropriate methodological behaviors by our fellow strategic management researchers, nor are we singling out any particular article for scrutiny. A large majority of the 69 papers in the sample would be susceptible to the same issues. Instead, we are simply attempting to parsimoniously demonstrate how the SMULs could potentially negatively impact theoretical conclusions and recommendations.

**Myth A**

Myth A, that coefficients on first-order terms are meaningless in models containing the multiplicative interaction term, seems to be commonly accepted. In 54 percent of the total sample (37 of 69) the authors offered no interpretation of the coefficients of the first-order terms in the full regression models that included the interaction terms, and instead focused on the coefficients in the base model only. This practice creates problems as discussed because the conditional nature of those relationships is ignored. The problem appears to be slightly less evident in the later period, with 31 percent of articles (4 of 13) published between 2010-2014 failing to interpret main effect coefficients in the full model, compared to 59 percent (33 of 56) of articles published between 2000 and 2009.

**Myth C**

Myth C, the practice of testing moderation models hierarchically, was evident in 58 of 69 (84 percent) of the articles in our total sample. This practice is especially common in the latter years. Every one of the thirteen articles in the 2010-2014 timeframe reported results using hierarchical regression models, compared to 80 percent (45 of 56) of articles published between 2000 and 2009. Thus, testing moderation using the hierarchical approach seems to be standard practice despite the known issues in doing so (see above).

Myths A and C pertain to the interpretation of coefficients. Interpreting direct or main effect coefficients in isolation from interaction terms leads to binary conclusions along the lines of “*X* is significantly related to *Y*” when the only defensible conclusion may be “*X* is significantly related to *Y* for particular values of *Z*”. Also, interpreting direct effect coefficients in a base model which excludes the theoretically relevant interaction term creates the problem of potentially biased coefficients due to omitted variables. The pervasiveness of hierarchical regression procedure creates conditions where it is all too easy for authors, reviewers, and readers alike to draw incorrect conclusions. We provide the following single example as an illustration of the potential interpretation problems which arise due to these pervasive myths.

As a part of their theoretical model, Haas and Hansen (2007) propose a relationship between an advisor’s lack of effort and the signaling of competencies. Specifically they pose the following hypothesis:

“*Hypothesis 3: Advisors’ experience improves the signaling of competencies, while their lack of effort decreases it*”. (Page 1140)

To test this hypothesis, the authors used the direct effect coefficient on the variable representing the advisor’s lack of effort in their Model 6 to conclude that there is no significant relationship between advisor’s lack of effort and the signaling of competencies. The relevant text from the article is quoted as follows.

*"Model 6 presents the effects of the knowledge content and process variables on the team’s ability to signal competence to the prospective client…However, teams that enlisted help from more experienced colleagues were able to signal their competence to their prospective clients more effectively, as predicted by Hypothesis 3. A lack of effort exerted by colleagues who advised the team had no significant effect, though, on the team’s ability to signal competence to the client. Hypothesis 3 therefore is only partially supported by the results shown in Model 6, since a team’s ability to signal competence was influenced by the content but not the process dimension of personal advice usage.” (Page 1148)* [emphasis added]

However the authors also suggest that the relationship proposed in Hypothesis 3 is moderated by advisor response time. They use the interaction term between advisor’s lack of effort and advisor response time in their Model 7 to conclude that there is a significant moderation effect which modified the relationship between advisor’s lack of effort and the signaling of competencies. We again quote from the original paper:

*“To further test for any possible effects of lack of effort by advisors, we examined whether the effects depended on advisor response time, and found that the interaction term between the effort and response time variables was significant and negative, as shown in Model 7. With an increase in response time, an increase in advisors’ lack of effort had a negative impact on the team’s ability to signal competence to the client. Stated differently, an increase in advisors’ effort had a more positive effect on competence signaling when response time was higher than when it was lower”. (Page 1149)*

This widely accepted practice of interpreting coefficients in a piecemeal fashion across a number of incomplete empirical models clearly causes a problem in the form of internally inconsistent conclusions. On Page 1148, Haas and Hansen (2007) draw an unconditional conclusion of “no significant effect” between advisors’ lack of effort and signaling of competencies. Based on this statement, future researchers who endeavor to build on their work may begin with the errant assumption that the theorized relationship between lack of effort and signal of competence does not actually exist. However, on the next page they conclude that there is in fact a relationship between those two variables under certain conditions related to response time. Rather than not supporting that portion of Hypothesis 3, their empirical tests actually offer qualified support. Advisor’s lack of effort does have a significant effect, sometimes. We draw an interaction plot to further represent the relevant relationships graphically (please see Figure 1), and find that when advisor response times are high, lack of effort has a significant negative effect on the signal of competence.

This example is indicative of the kinds of problems created by hierarchical regression models. It is exceptionally easy for authors and readers to draw inaccurate and overly simplistic conclusions based on base models, when the reality of the situation is more nuanced. A further reflection tells us the cause of the problem begins with the hypothesis development itself. It is inconsistent to offer a binary direct effect hypothesis like in Hypothesis 3, when other parts of the theoretical model suggest that “it depends”. It is better to hypothesize the direction and size of the effects across the ranges of the variables involved instead, for example “Advisor’s lack of effort decreases the signaling of competencies when the advisor takes a long time to respond”.

-----------------------------------------

Insert Figure 1 about Here

-----------------------------------------

**Myth B**

Myth B is very common: Reporting the reliability of the interaction term was almost universally absent, with only 2 of 69 articles published between 2000 and 2009 doing so. The others (67 of 69, 97%) either report Cronbach's alpha only for the first-order terms or do not address variable reliability at all. The omission of measurement reliability of the interaction term is noteworthy since a conventionally acceptable level of reliability for the first-order terms is no guarantee of acceptable reliability for their product.

Again, we only provide one example to illustrate how low reliability of interaction term could influence interpretation of results. Shaw, Gupta and Delery (2001) examined interactive relationships between the use of integrated manufacturing and compensation practices in predicting plant performance. They hypothesized that advanced manufacturing technology (AMT) and total quality management (TQM) can combine with team compensation incentives and individual compensation incentives to help explain workforce effectiveness (Hypotheses 1 & 2). Based on findings from hierarchical linear regressions, the authors conclude “There was virtually no support for the interactions including AMT” and “Hypothesis 1, interactions with individual incentives, was the most consistently supported, but only with TQM… Only very spotty support for Hypotheses 2 (and 3) was found” (page 382). The authors interpreted:

*“While interesting, the lack of uniform support may indicate that the predicted relationships fail to completely capture the dynamics of manufacturing system-HR system fit, that we lacked sufficient power to detect more consistent effects, and/or that measurement limitations diminished the sensitivity of our analyses. These limitations suggest that the hypotheses may not be fully tested in our study.” (Page 382)*

However, measurement error of interaction terms could be another plausible explanation. We calculate the reliability of interaction terms in their study based on equation (2). The reliability for ATM, TQM, individual incentive, and team incentive variables were 0.78, 0.87, 0.72 and 0.74 respectively (page 383). Taking into account the correlations among the main effect variables, the reliability values of the four interaction terms (AMT × Individual incentives, AMT × Team incentives, TQM × Individual incentives, TQM × Team incentives) were 0.57, 0.58, 0.64, and 0.68, respectively. Clearly, the reliability values of interaction terms related to AMT are only 0.57 and 0.58 which do not meet even the mythical .70 standard commonly attributed to Nunnally (see footnote 3). In comparison, the reliability values of the interaction terms related to TQM are 0.64 and 0.68, which are approaching the mythical .70 standard. Measurement reliability can affect power via attenuation which is a statistical concept that refers to underestimating the correlation between two different measures because of measurement error (Boyd, et al., 2005a). This seems to offer another plausible explanation of the inconsistent moderation findings related to AMT in addition to the two reasons the authors offer – measurement limitations (AMT focuses simply on technological aspects of manufacturing, may have less-ranging impact and limited benefit) and lack of power (due to small sample size).

**TOWARD BETTER PRACTICES**

When conducting moderation analyses many strategic management scholars appear to have relied on statistical myths and urban legends when conducting, reporting, and interpreting their findings. Methodological precedence may have been driven more by what previous strategic management researchers have done rather than the application of sound statistical practices. And more importantly, these practices can (and apparently have) lead to errors in empirical conclusions, potentially leading to mistaken interpretations and incorrect recommendations. Our results highlight three likely areas for concern.

First, misinterpretation can be reduced if researchers abandon the hierarchical regression paradigm. Currently, most strategic management researchers present a model of main effects, another with the interaction terms, and then examine the change in model fit (*R*2). However, a full regression model containing all theoretically relevant variables and controls will lead to less biased coefficients compared to base models that omit the interactions. In addition, *t*-tests on the interaction coefficients tell us just as much about the significance of the moderation effect as *F*-tests in hierarchical models (J.R. Edwards, 2009). Further, we suggest that researchers stop the practice of interpreting either the direct effect coefficients or the interaction coefficients in isolation. By their very nature, moderation models suggest that the effect of one variable will depend on the level of another, and therefore a full and correct interpretation can only be achieved by considering all of those coefficients jointly. We recommend that researchers interpret the main effects only when moderating effects are non-significant; otherwise, researchers expose their findings to both Type I and II errors (Carte & Russell, 2003). We also must exercise caution when interpreting direct effects, even in the full model. As Edwards (2009) points out, when the underlying theoretical model being tested has a moderated form, the effects that the independent variables and the moderators have on the dependent variable can only be described as being conditional. In other words, when the theoretical model itself suggests that the effect of *X* on *Y* depends on the value of *Z*, it is incomplete and potentially incorrect to make any binary statements based on a point estimate of a direct effect coefficient. Further it is important to draw interaction plots to aid demonstration and interpretation of moderation effect in addition to running multiple moderated regression analysis (Aiken & West, 1991).

This recommendation extends to the format commonly used to formulate hypotheses around moderation models. The typical practice of hypothesizing that “*X* positively/negatively affects *Y*” followed by a separate hypothesis that “The strength of the *X-Y* relationship depends on *Z*” can, in many cases, lead to misleading conclusions. If the theoretical model in question suggests moderation in which the relationship between *X* and *Y* depends on *Z*, it is inconsistent to offer a binary *X-Y* direct effect hypothesis without, at the very least, specifying the value of *Z* for which the authors expect that relationship to hold. Authors need to be explicit in their expectations regarding the direction and size of the effects across the ranges of the variables involved.

Second, even highly reliable direct effect variables can yield interaction terms which fail conventional standards for measurement error. Measurement error can make the detection and interpretation of a significant interaction effect problematic and could lead to different conclusions than otherwise would not be reached (e.g., Type II errors and an exaggeration of *R*2 values). Authors need to be aware of this potential problem and report the reliability of interaction terms. At present, relying simply on precedent, or hunting for other studies that might have used the same rules, can only serve to perpetuate problems in how studies are done and interpreted. Considering that some of the studies in our analysis have interaction terms that have reliabilities in the range of .5, the conclusions and recommendations based upon those variables may be influenced by error. For example, a reliability of .5 means that 50% of the variance is due to error, and 50% due to the true score. Such situations are clearly inconsistent with the field’s interests in raising the quality of its science and the validity of its conclusions. We therefore recommend researchers to be mindful about reliability of interaction terms and apply equation (2) to understand the level of reliability of interaction terms no matter what levels of reliability the X and Z variables have.

To examine how the reliabilities of interaction terms would change if reliability of main effect variables were different, we conducted other calculations using equation (2) by manipulating levels of reliability of main effect variables of two studies. We found that when the reliability for main effect variables is above .8, the reliability of interaction terms are around .7. When the reliability of X and Z both exceed .9, then the reliability of interaction terms reaches around .8. And when the reliability of the main effects are below .8, then the reliability of the interaction terms will tend to fall below even the mythical standard of .7 (see footnote 3; Lance et al., 2006). Obviously, the results are influenced by the level of correlation between the two main effect variables as well. We are aware that some researchers argue that an excessively high reliability is not desirable, since that would suggest that the measurements contain repetitive and overlapping items, and as such, that the value of Cronbach’s Alpha should not exceed .9 (e.g., Streiner, 2003). Also, what should be considered an acceptable level of reliability depends on the stage of development of the measure (Lance et al., 2006). Hence, we are not recommending mechanistically applying any particular cut-off criterion for reliability. Rather we are recommending transparency and disclosure with respect to the selected reliability values so that reviewers and readers may draw their own conclusions.

Related to Myth B, one might ask what if the moderator is a dummy or categorical variable. In this case, the reliability of the moderating variable could be measured by other reliability tests other than internal consistency tests, such as inter-rater agreement Cohen’s kappa (Cohen, 1960) in OB research. However, in strategic research, reliability of these measurements is rarely reported. In our sample, only one paper reported Cohen’s kappa. Reliability and measurement error is rather inherent to multiple items measurements in strategic research (Boyd, et al., 2005b). Indeed, in our sample, 16 papers reported Cronbach’s alpha and two papers reported composite scale reliabilities of multiple item measurements. This represents about 26% of our sample, quite comparable to the 33.3% reported in Boyd, et al. (2005b). Taken together, the majority of strategic research do not report reliabilities. Future research in strategic management needs to take more care of reliability and measurement error of variables, interaction terms, and other transformed variables. We also found in our sample of 69 papers that some authors dichotomized originally and conceptually continuous moderating variables. Several authors have recommended against this practice (e.g., Aguinis & Gottfredson, 2010; Dawson, 2014) since dichotomization and similar conversions of the original continuous measures lead to the introduction of nonlinear and nonrandom measurement error (Maxwell & Delaney, 1993) and substantial power loss in tests of moderators (Aiken & West, 1991; Stone-Romero & Anderson, 1994). Hence future strategic research should avoid the practice of converting conceptually justified continuous variables to categorical variables.

Finally, we suggest that strategic management researchers adopt a “best practices” approach to testing moderation analysis rather than relying on what other strategists have done as the basis for their statistical practices (e.g., Aguinis & Gottfredson, 2010; Carte & Russell, 2003; Murphy & Russell, 2017). Indeed, methodological advances in the strategic management field often take the form of redirecting problematic practices (Bascle, 2008; Bergh & Fairbank, 2002; Hamilton & Nickerson, 2003). Thus, the current practice of basing methodological decisions on previous research, which may lead to the propagation of problematic practices, should be re-examined. This recommendation extends beyond moderation. Other SMULs might exist within the field, and they could similarly be impacting our results and conclusions. Indeed, very little research exists about how decisions about such things as samples, outliers, and missing values are handled and as such, researchers are left to their own devices to navigate such situations. This unchartered territory is potentially problematic, as SMULs appear to be widespread (e.g., Lance and Vandenberg, 2009, 2015).

The foregoing recommendations are essential for reversing what have become conventional practices that can have negative implications on knowledge accumulation and recommendations for practice. Our findings suggest that conventional standards - those that serve as precedents within the field - may be based on myths and urban legends. To the extent that those myths materially affect the conclusions we draw from our research, as illustrated here, the recommendations we make to practitioners guided by those research findings might not be accurate. SMULs are not visible to corporate executives but could have a profound impact on their strategic decision making indirectly through the knowledge created by strategic research. If the field of strategic management is to advance and provide a compelling basis for informing decision making, reconsidering how moderation analysis and methodological decision making in general is conducted and interpreted appears to be warranted.

We echo Starbuck’s view that “the current culture of social science research supports defective norms that resist reform” (Starbuck, 2016, page 9). Since some SMULs have become accepted conventions in our research methodology, papers that follow the mythological precedents get published in highly ranked journals and those that do not may not get published at all. These top journals also set standards for training of researchers and PhD students, hence these conventional practices are further reinforced. In order for strategic management research specifically and social science research at large to embrace reforms, abandoning SMULs is a necessary step to take. Based on the findings from this study, we strongly recommend editors and reviewers of journals to require authors not report direct effect only models, rather report full moderation models, and also report reliability of interaction terms.

**CONCLUSION**

In conclusion, potentially problematic practices have become gold standards within strategic management research and may have adversely influenced the interpretation and reporting of moderation tests, led to incomplete findings, errant conclusions and even incorrect interpretations. We hope that our study helps bring these decisions and their respective solutions to light and encourages researchers henceforth to reconsider the conventional processes that seem to have become institutionalized within the field and to adopt rigorous approaches to justifying methodological decisions. Ultimately, the confidence we place in research findings and the sustainability of knowledge and theoretical explanations depends on sound practices and standards in the field.

**REFERENCES**

**Aguinis, H. and R. K. Gottfredson,** 2010, Best practice recommendations for estimating interaction effects using moderated multiple regression. *Journal of Organizational Behavior,* **31:** 776-786.

**Aguinis, H. and E. F. Stone-Romero,** 1997, Methodological Artifacts in Moderated Multiple Regression and Their Effects on Statistical Power. *Journal of Applied Psychology,* **82:** 192-206.

**Aiken, L.S. and S.G. West,** 1991**,** *Multiple regression: testing and interpreting interactions*. Sage.

**Arnold, H. J. and M. G. Evans,** 1979, Testing multiplicative models does not require ratio scales. *Organizational Behavior and Human Performance,* **24:** 41-59

**Bascle, G., 2008,** Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization*, **6:** 285-327.

**Bergh, D.D., and J. Fairbank,** 2002, The measurement and testing of change in strategic management research. *Strategic Management Journal*, **23:** 359-366.

**Bergh, D. D., B. M. Sharp, H. Aguinis, and M. Li,** 2017**,** Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. *Strategic Organization,* **15:** 423-436.

**Bohrnstedt, G. W. and G. Marwell,** 1978, The Reliability of Products of Two Random Variables. *Sociological Methodology,* **9:** 254-273.

**Boyd, B. K., D. D. Bergh, R. D. Ireland,** **and D. J. Ketchen,** 2013, Constructs in Strategic Management. *Organizational Research Methods,* **16:** 3-14.

**Boyd, B. K., S. Gove, and M. A. Hitt,** 2005a**,** Consequences of measurement problems in strategic management research: the case of Amihud and Lev. *Strategic Management Journal,* **26:** 367-375.

**Boyd, B. K., S. Gove, and M. A. Hitt,** 2005b**,** Construct measurement in strategic management research: illusion or reality? *Strategic Management Journal,* **26:** 239-257.

**Boyd, B. K., K. T. Haynes, M. A. Hitt, D. D. Bergh, and D. J. Ketchen,** 2012, Contingency hypotheses in strategic management research. Use, disuse or misuse? *Journal of Management*, **38:** 278-313.

**Carte, T. A., and C. R. Russell,** 2003, In pursuit of moderation: nine common errors and their solutions. *MIS Quarterly,* **27:** 479-501.

**Champoux, J. E. and W. S. Peters,** 1987, Form, effect size and power in moderated regression analysis. *Journal of Occupational Psychology,* **60:** 243-255.

**Cohen, J.,** 1960, A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement,* **20:** 37-46.

**Cohen, J.,** 1978, Partialed products are interactions; partialed powers are curve components. *Psychological Bulletin,* **85:** 858-866.

**Cortina, J. M.,** 1993, Interaction, Nonlinearity, and Multicollinearity: implications for Multiple Regression. *Journal of Management,* **19:** 915-922.

**Cortina, J. M., G. Chen, and W. P. Dunlap,** 2001, Testing Interaction Effects in LISREL: Examination and Illustration of Available Procedures. *Organizational Research Methods,* **4:** 324-360

**Cronbach, L. J.,** 1987, Statistical tests for moderation variables: Flaws in analyses recently proposed. *Psychological Bulletin*, **102:** 414-417,

**Dalal, D. K. and M. J. Zickar,** 2012, Some common myths about centering predictor variables in moderated multiple regression and polynomial regression. *Organizational Research Methods***, 15:** 339-362.

**Dawson, J.,** 2014, Moderation in Management Research: What, Why, When, and How. *Journal of business and Psychology*, **29**(1): 1-19.

**Dunlap, W. P. and E. R. Kemery,** 1987, Failure to detect moderating effects: Is multicollinearity the problem? *Psychological Bulletin,* **102:** 418-420.

**Dunlap, W. P. and E. R. Kemery,** 1988, Effects of predictor intercorrelations and reliabilities on moderated multiple regression. *Organizational Behavior and Human Decision Processes,* **41:** 248-258.

**Durand, R., R. M. Grant, and T. L. Madsen,** 2017, The expanding domain of strategic management research and the quest for integration. *Strategic Management Journal,* **38:** 4-16

**Echambadi, R. and J. D. Hess,** 2007, Mean-Centering Does Not Alleviate Collinearity Problems in Moderated Multiple Regression Models. *Marketing Science,* **26:** 438-445.

**Edwards, J. and L. S. Lambert,** 2007, Methods for Integrating Moderation and Mediation: A General Analytical Framework Using Moderated Path Analysis. *Psychological Methods,* **12:** 1-22.

**Edwards, J. R.,** 2001, Multidimensional Constructs in Organizational Behavior Research: An Integrative Analytical Framework. *Organizational Research Methods***, 4:** 144-192.

**Edwards, J. R.,** 2009, Seven deadly myths of testing moderation in organizational research. In CE Lance and RJ. Vandenberg (eds.), *Statistical and Methodological Myths and Urban Legends: Received Doctrine, Verity, and Fable in the Organizational and Social Sciences* (pp. 143-164). New York, NY: Routledge.

**Ganzach, Y.,** 1998, Nonlinearity, Multicollinearity and the Probability of Type II Error in Detecting Interaction. *Journal of Management,* **24:** 615-622.

**Hamilton, B. H., and J. A. Nickerson,** 2003, Correcting for endogeneity in strategic management research. *Strategic Organization*, **1:** 53-78.

**Haas, M. R. and M. T.** **Hansen,** 2007, Different knowledge, different benefits: toward a productivity perspective on knowledge sharing in organizations. *Strategic Management Journal,* **28**: 1133-1153.

**Jaccard, J. and C. K. Wan,** 1995, Measurement error in the analysis of interaction effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches. *Psychological Bulletin,* **117:** 348-357.

**Jaccard, J., C. K. Wan, and R. Turrisi,** 1990, The Detection and Interpretation of Interaction Effects Between Continuous Variables in Multiple Regression. *Multivariate Behavioral Research,* **25:** 467-478.

**Kenny, D. A. and C. M. Judd,** 1984, Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin,* **96:** 201-210.

**Kromrey, J. D. and L. Foster-Johnson,** 1998, Mean Centering in Moderated Multiple Regression: Much Ado about Nothing. *Educational and Psychological Measurement,* **58:** 42-67.

**Lance, C. E.,** 2011, More statistical myths and urban legends. *Organizational Research Methods,* **14:** 279-286.

**Lance, C. E., M. B. Butts, and L. C. Michaels,** 2006, The sources of four commonly reported cutoff criteria: what did they really say? *Organizational Research Methods*, **9:** 202-220.

**Lance, C. E., and R. J. Vandenberg,** 2009, *Statistical and Methodological Myths and Urban Legends: Received Doctrine, Verity, and Fable in the Organizational and Social Sciences*. New York, NY: Routledge.

**Lance, C. E., and R. J. Vandenberg,** 2015, *More Statistical and Methodological Myths and Urban Legends: Received Doctrine, Verity, and Fable in the Organizational and Social Sciences*. New York, NY: Routledge.

**Li, F., P. Harmer, T. E. Duncan, S. C. Duncan, A. Acock, and S. Boles,** 1998, Approaches to Testing Interaction Effects Using Structural Equation Modeling Methodology. *Multivariate Behavioral Research,* **33:** 1-39.

**MacCallum, R. C. and C. M. Mar,** 1995, Distinguishing between moderator and quadratic effects in multiple regression. *Psychological Bulletin,* **118:** 405-421.

**Maxwell, S. E. and H. D. Delaney,** 1993, Bivariate median splits and spurious statistical significance. *Psychological Bulletin*, **113:** 181-190.

**McClelland, G. H. and C. M. Judd,** 1993, Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin,* **114:** 376-390.

**Mizruchi, M., and L. Fein,** 1999, The social construction of organizational knowledge. 1999, *Administrative Science Quarterly*, **44:** 653-683

**Murphy, K. R. and C. J. Russell,** 2017, Mend It or End It: Redirecting the Search for Interactions in the Organizational Sciences. *Organizational Research Methods,* **20:** 549-573.

**Nag, R., D. C. Hambrick, and M-J Chen,** 2007, What is strategic management, really? Inductive derivation of a consensus definition of the field. *Strategic Management Journal,* **28:** 935-955.

**Nunnally, J. C.,** 1978, *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

**Ping, R. A., Jr.,** 1996, Latent variable interaction and quadratic effect estimation: A two-step technique using structural equation analysis. *Psychological Bulletin,* **119:** 166-175.

**Preacher, K. J., Z. Zhang, and M. J. Zyphur,** 2016, Multilevel structural equation models for assessing moderation within and across levels of analysis. *Psychological Methods,* **21:** 189-205.

**Shaw, J. D., N. Gupta, and J. E. Delery,** 2001, Congruence between technology and compensation systems: implications for strategy implementation. *Strategic Management Journal*, **22:** 379-386.

**Shook, C. L., D. J. Ketchen Jr, C. S. Cycyota, and D. Crockett,** 2003, Data analytic trends and training in strategic management. *Strategic Management Journal,* **24:** 1231-1237.

**Spector, P. E.,** 2006, Method variance in organizational research: Truth or urban legend?

*Organizational Research Methods,* **9:** 221-232.

**Starbuck, W. H.,** 2016, 60th anniversary essay: How journals could improve research practices in social science. *Administrative Science Quarterly,* **61:** 165-183.

**Stone-Romero, E. F. and L. E. Anderson,** 1994, Relative Power of Moderated Multiple Regression and the Comparison of Subgroup Correlation Coefficients for Detecting Moderating Effects. *Journal of Applied Psychology,* **79:** 354-359.

**Streiner, D. L.,** 2003, Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment,* **80:** 99-103.

**Vandenberg, R. J.,** 2006, Statistical and methodological myths and urban legends: Where pray tell did they get this idea? *Organizational Research Methods,* **9:** 194-201.

**Venkatraman, N.,** 1989, The concept of fit in strategy research : Towards verbal and statistical correspondence. *Academy of Management Review,* **14:** 423-444.

**Williams, L. J., R. J. Vandenberg, and J. R. Edwards,** 2009, Structural equation modeling in management research: A guide for improved analysis. *The Academy of Management Annals,* **3:** 543-604.

**Figure 1: Derived from Haas and Hansen (2007), showing the effect of lack of effort on signal of competence, moderated by response time**

**Table 1. Seven SMULs associated with testing moderation (Edward, 2009)**

|  |  |  |  |
| --- | --- | --- | --- |
| SMULS | Typical procedure | Recommendations by Edward (2009) | Most significant references |
| Myth 1: Interaction terms create multicollinearity problems | Mean-centering of the terms used to create the interaction | No mean-centering; the correlation of interest concerns the independent variables with one another, not with the cross-product term | Echambadi & Hess (2007);  Dalal & Zickar (2012);  Cohen (1978);  Cronbach (1987);  Dunlap and Kemery (1988);  Kromrey and Foster-Johnson (1998) |
| Myth 2 (Myth A): Coefficients on first-order terms are meaningless | When the interaction term is in the equation, then the coefficients for the first-order (main effect) variables are inconsequential | These variables provide valuable information and can help produce a more comprehensive understanding of the data; they need to have means within the range of the data and can be used to compute simple slopes of the interaction terms | Aiken & West (1991);  Jaccard, Wan, and Turrisi (1990);  Champoux & Peters (1987) |
| Myth 3 (Myth B): Measurement error poses little concern when first-order terms are reliable | When the first-term (main effect) variables have acceptable reliability, researchers ignore the measurement properties of the interaction term | Report the reliability of the interaction term (may be influenced by the variable scale’s; consult Edwards, 2009) | Jaccard & Wan (1995);  Aguinis & Stone-Romero (1997);  Arnold and Evans (1979);  Cohen (1978);  Dunlap and Kemery (1987) |
| Myth 4 (Myth C): Interaction terms should be tested hierarchically | Researchers test moderation in a two-step process, where main effects are considered in step 1 and interaction term in step 2 | Test main effects and interaction terms in same model | Cohen (1978);  Jaccard, et al. (1990);  McClelland and Judd (1993);  Kromrey and Foster-Johnson (1998);  Cronbach (1987); |
| Myth 5: Curvilinearity can be disregarded when testing moderation | Researchers ignore curvilinearity, except when it is part of theoretical model | Include curvilinearity tests, as they provide a stronger test of the hypothesis | Cortina (1993);  Ganzach (1998);  MacCallum & Mar (1995);  Dawson (2014); |
| Myth 6: Interaction terms can be treated as causal variables | Researchers use the interaction term as component of mediation tests and assume that it can assume a causal format | The interaction term has no causal properties, as it is merely a mathematically-derived variable with no unique properties of its own | Edwards & Lambert (2007) |
| Myth 7: Testing moderation in structural equations modeling is impractical | Most moderation tests using ANOVA and regression analysis; vulnerable to problems associated with measurement error | The application of structural equations modeling is becoming less complex and the advantages of using it offsets the detriments associated with measurement problems | Cortina, Chen, & Dunlap (2001);  Kenny & Judd (1984);  Jaccard & Wan (1995);  Preacher, Zhang, & Zyphur (2016);  Li, Harmer, Duncan, Duncan, Acock, & Boles (1998);  Ping (1996)  Williams, Vandenberg, and Edwards (2009) |

**APPENDIX**

The coding scheme operationalizes Myths 2, 3 and 4 (Edward, 2009). We deconstructed each myth description into its detailed components, created variables for each, developed coding questions for each variable in an explicit and clear manner, and added open-ended questions to capture additional textual-related details.

The coding scheme was first applied to two articles, each of which was coded independently by two of the authors. Comparisons of the initial coding values revealed an 88 percent agreement rate. Differences were discussed and the coding questions were revised and clarified until consensus was reached on those two articles. Next, to verify that the newly revised coding scheme would provide consistent independent codings, the comparison process was repeated for three additional articles. The process of discussion, revision, and clarification was repeated until consensus was reached on those three articles and all authors were confident that the coding scheme was valid and could be applied in a reliable fashion. Afterwards, the remaining moderation articles were divided between the coders for independent classification. At the end of the coding process for the moderation articles, the two coders independently coded and compared two articles that had been assigned to the other coder. In addition, to test for generalizability, four additional articles were randomly selected from the population of non-selected *SMJ* moderation articles and coded. The post-hoc reliability tests across these six articles revealed no differences between the coders. Table below reports the coding scheme and results.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Myth** | **Variable** | **Coding question** | **Code** | **Number of articles** | | | | | |
|  |  | Yes/1 | | No/0 | | | Missing |
|  | 2000-2009 | 2010-2014 | 2000-2009 | | 2010-2014 |  |
| 2 (A) | MOD2.1 | Did the authors interpret the coefficients of the first-order terms in the base model? | 0 = No 1 = Yes | 39  70% | 7  54% | 16  29% | | 6  46% | 1  2% |
| 2 (A) | MOD2.2 | Did the authors interpret the coefficients of the first-order terms in the models with interaction terms? | 0 = No 1 = Yes | 22  39% | 9  69% | 33  59% | | 4  31% | 1  2% |
| 2 (A) | MOD2.3 | Did the authors state that the coefficients of the first-order terms are meaningless WHEN testing a moderation relationship? | 0 = No 1 = Yes | 0 | 0 | 56  100% | | 13  100% | 0 |
| 3 (B) | MOD3.1 | Did the authors interpret the measurement errors of the first-order terms? | 0 = No 1 = Yes | 3  5% | 6  46% | 53  95% | | 7  54% | 0 |
| 3 (B) | MOD3.2 | Did the authors interpret the measurement errors of interaction terms? | 0 = No 1 = Yes | 2  4% | 0 | 54  96% | | 13  100% | 0 |
| 3 (B) | MOD3.3 | Did the authors state that the measurement errors is of little concern when first order terms are reliable? | 0 = No 1 = Yes | 1  2% | 0 | 55  98% | | 13  100% | 0 |
| 3 (B) | MOD3.4 | Did the authors report Cronbach's alpha to support the reliability of the first-order terms? | 0 = No 1 = Yes | 12  21% | 4  31% | 44  79% | | 9  69% | 0 |
| 3 (B) | MOD3.5 | If MOD3.4 = 1, what was the cutoff Cronbach's alpha value the authors considered acceptable? | Open | Only two articles gave an explicit cutoff value, both used 0.7 | | | | | |
| 3 (B) | MOD3.6 | If the authors reported reliability but not Cronbach's alpha value, what measure did they use? | Open | Only three papers used an alternative reliability measure. They reported “composite scale reliability”, “inter-rater reliability”, and “Cohen's kappa”. | | | | | |
| 3 (B) | MOD3.7 | If the authors reported reliability but not Cronbach's alpha value, what cutoff reliability value did they use for the measure they adopted? | Open | One paper used a cutoff of 0.6 for their composite scale reliability | | | | | |
| 4 (C) | MOD4.1 | Did the authors test the interaction term XZ hierarchically: First estimating an equation using only X and Z as predictors, and then estimating an equation that adds XZ? | 0 = No 1 = Yes | 45  80% | 13  100% | 11  20% | 0 | | 0 |
| 4 (C) | MOD4.2 | If MOD4.1=1, did the authors test the difference of the two equations using the F-ratio or its equivalent? | 0 = No 1 = Yes | 27  48% | 7  54% | 17  30% | 6  46% | | 12  21% |
| 4 (C) | MOD4.3 | If MOD4.1 = 1, did the authors cite previous literature as support for testing the interaction terms hierarchically? | 0 = No 1 = Yes | 2  4% | 0 | 42  75% | 13  100% | | 12  21% |

1. Statistical myths and urban legends are apparent in many methodological decisions. See Lance and Vandenberg (2009, 2015) for applications. [↑](#footnote-ref-1)
2. See also Aguinis and Gottfredson (2010) for pre-data collection and post-data collection recommendations, as well as Carte and Russell (2003) for guidelines on measurement decisions and reporting disclosures. [↑](#footnote-ref-2)
3. Nunnally (1978) thoughtfully addressed satisfactory levels of reliability in various situations, saying that what is considered adequate can range from 0.7 to 0.95. His recommendation was that “In the early stages of research…one saves time and energy by working with instruments that have only modest reliability, for which purpose reliability of .70 or higher will suffice…” (page 245). Following Nunnally, once research advances and instruments become more established, the bar for reliability should increase. It is also worth noting what many researchers claim implicitly or explicitly about a cut-off reliability value of .70 based on Nunnally (1978) is a myth based on an incomplete understanding of the author’s original intent. [↑](#footnote-ref-3)