

---

# Narrow Standards for Efficacy and the Research Playground: Why Either–Or Conclusions Do Not Help

---

ROBERT G. JONES  
*Missouri State University*

RICHARD J. KLIMOSKI  
*George Mason University*

Lance's (2008) article on assessment center (AC) construct validation has undertones of the long-standing "lumper" versus "splitter" discussions in biology. The controversy here is between breaking groups of animals into more categories (i.e., subspecies, subfamilies, etc.) or whether broader "lumping" of categories together is more accurate. Lance argues for "splitting" AC results into behavioral and "role" dimensions within exercises, while behavior reporting (across-exercise dimension ratings) relies on fewer rating categories that attempt to cover broader patterns of behavior. Although both lumpers and splitters are attempting to accomplish the same outcome—to simplify and categorize observed phenomena into parsimonious categories—such controversies do little to advance emergent understanding of phenomena. The same can be said of Lance's attempt to dismiss one type of AC rating category for another.

Lance concludes that, because predictive accuracy is not affected by AC design, we should scrap the more cumbersome behavioral reporting (dimension-based) method. We see two major problems with this. If psychometric evidence is to be our only (narrow) standard for adopting AC features, we argue that this conclusion is premature. There is simply not enough comparative data. Furthermore, Lance appears to adopt a very narrow perspective in making some very sweeping recommendations. In doing this, we feel that he ignores the multiple functions that the AC method fulfills in actual operational settings and the multiple reasons that this very expensive, complex human resource (HR) management practice has been sustained over the years.

## Applying the Narrow Standard

Thornton and Rupp (2006) provided a description of the many, varied uses of the AC method. These include personnel selection and support for promotion decisions, for which magnitude of the correlation between center ratings and various criteria represents an important technical standard. Similarly, the outcomes of multitrait–multimethod (MTMM) analyses provide an additional psychometric standard. For the scientist–practitioner, the

---

Correspondence concerning this article should be addressed to Robert G. Jones. E-mail: robertjones@missouristate.edu

Address: Department of Psychology, Missouri State University, Springfield, MO 65897.

Robert G. Jones, Department of Psychology, Missouri State University; Richard J. Klimoski, School of Management, George Mason University.

weighing of these standards to meet client needs would tend to argue for the former (prediction) over the latter (MTMM) standard. To our knowledge, there is no evidence regarding differential prediction yielded from exercise- versus dimension-based ratings. So, by the predictive standard in selection and promotion centers, there is little obvious reason for choosing one over the other method.

We doubt, however, whether many managers would argue that predictive accuracy should be the sole basis for determining the usefulness even of selection devices. Muchinsky (2004) has argued convincingly that selection devices are one method for accomplishing the broader purpose of organizational change. He states that “statistically based arguments do not address the full range of issues under consideration,” which also include cost per incremental improvement in prediction (Lance’s implicit standard, we think), perceived fairness, ethical use of data, adverse impact, security of individual scores, organizational acceptability, and effective downstream implementation. To illustrate these issues, Muchinsky goes as far back in the process of developing a selection device as the selection of subject matter experts.

With the AC method, we feel that there is a need to go even further “upstream” in our thinking to address the kind of organizational impact desired by implementing the method. If the prediction of performance is what a client organization wants, then perhaps the “job sample” method (within-exercise ratings) would be preferred (in terms of cost per increment in performance) over the behavioral reporting method. However, if other purposes are being served (e.g., to promote a more level playing field for workers seeking promotions, serving as the basis for developmental discussions), we have no evidence at all about what would constitute the appropriate AC format. Moreover, we do not suggest that ACs are either better or worse than other methods (e.g., multisource ratings) for some of these purposes. We simply argue that it is premature to arrive at conclusions about the “preferred” AC design.

## **Shared Assumptions and Different Conclusions**

We disagree with Lance’s “either-or” conclusion, even about the psychometric adequacy of behavioral reporting mostly based on a complete lack of comparative longitudinal predictive evidence (all longitudinal research has been done using behavioral reporting) or comparative MTMM analyses (all MTMM analyses have been done using within-exercise ratings). However, we do share his assumption that highly experienced and trained observers may be accurate judges of people and their potential. For example, we have argued elsewhere that assessors may be using a combination of inferences about individual differences (traits) and situational constraints to make their judgments. So, if we start from this assumption, different sorts of design and implementation scenarios other than within- or between-exercise approaches can be entertained. For example, rather than crafting an AC inductively (from a job analysis) or deductively (using prototypic dimensions), we might instead start with what the average assessor (manager) can do intuitively.

## **Mission Shift and Building a Better Mouse Trap**

Regardless of how interesting such arguments about assessor capabilities may be, current conjecture based on existing AC research leads us astray and might be thought of as an example of “mission shift,” where the means have become the ends. For many, the AC method was seen as a platform for making better HR decisions. But in the 1970s, many academics became enamored with the cognitive revolution and new developments in social-cognitive psychology. Along the way, the AC method became one venue for learning about human judgment and decision processes. Often there was only peripheral interest in informing AC design in operational settings. This, in turn, produced the literature associated with AC dimensional construct validity. In our view, over the years, the point of it all became

muddled. The means became the ends because researchers attempted to find ways to enhance the validity of dimensional judgments, rather than enhancing the value of ACs to organizations.

Indeed, some of us had hoped that research on cognitive processes uncovered over years of research on the method could be translated into promoting not just more construct-valid but better (more useful) assessments in applied settings. But in retrospect, it is not clear that we went about it in the right way. Contrary to Boehm's (1980) admonition, we did not approach our AC research with implementation in mind. To exaggerate a bit, the majority of the investigators of the method, in addressing the "meaning" behind MTMM analyses of within-exercise ratings, seemed only tangentially interested in improving practice. Moreover, implementation challenges were rarely considered when making research-based recommendations. In any event, proposed changes were rarely evaluated empirically in actual operational settings to see if they would make a meaningful difference to stakeholders.

We readily acknowledge that the AC method is a HR practice that should be informed by scientific findings. But it must be a method that is not just professionally defensible (e.g., have predictive and/or content-construct validity). To be useful and to be institutionally sustainable, any revised method approach must also be feasible (practical) and acceptable to key stakeholders (Austin, Klimoski, & Hunt, 1996). The advice offered by Lance needs to be reconsidered relative to these additional criteria.

## Conclusions

If our sole purpose in studying the AC was to understand the social psychology of group decision making or the cognitive processes

that lead assessors to arrive at high predictive accuracy, we certainly have learned a lot. Yet, ACs are not just research sites for investigators. They are typically implemented by work organizations at considerable expense in order to support HR goals, only one of which may be to assist in the prediction of future performance. As scientist-practitioners, then, we should frame our investigations with such a diversity of purposes kept in mind.

Lance or others who advocate a within-exercise approach to AC design, with all of the developmental work and managerial oversight that this would involve, have an obligation to clearly establish its "value proposition" in light of the firm's needs or goals. To address the question "Is the within-exercise approach better?", we would also need to know the answer to such additional questions as "for whom," "for what purpose," "compared to what," "at what cost," and "for how long?" But, even if we treat the AC method as our research playground, there are plenty of questions and assumptions that require further evaluation before arriving at either-or conclusions.

## References

- Austin, J. T., Klimoski, R. J., & Hunt, S. T. (1996). Dilemmas in public sector assessments: A framework for developing and evaluating selection systems. *Human Performance, 9*, 177-198.
- Boehm, V. R. (1980). Research in the "real world": A conceptual model. *Personnel Psychology, 33*, 495-503.
- Lance, C. E. (2008). Why assessment centers do not work the way they are supposed to. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 84-97.
- Muchinsky, P. M. (2004). When the psychometrics of test development meets organizational realities: A conceptual framework for organizational change, examples, and recommendations. *Personnel Psychology, 57*, 175-209.
- Thornton, G. C., & Rupp, D. E. (2006). *Assessment centers in human resource management: Strategies for prediction, diagnosis, and development*. Mahwah, NJ: Erlbaum.