

# It Is Not Yet Time to Dismiss Dimensions in Assessment Centers

KLAUS G. MELCHERS AND CORNELIUS J. KÖNIG  
*Universität Zürich*

Past research has well documented that assessors' postexercise dimension ratings (PEDRs) in an assessment center (AC) reflect the different exercises much more than the different dimensions. This is often understood as a problem for the construct validity of ACs. However, Lance (2008) does not consider the construct validity of ACs to be troubling. On the basis of his description of the situation (which we consider to be very accurate), Lance instead either explicitly or implicitly makes the following suggestions: (a) regard ACs as a collection of work sample tests, (b) treat only exercise variance in assessors' PEDRs as relevant variance, and (c) give up the search for design fixes that might improve construct validity. We would like to challenge these three suggestions.

## **Does Considering ACs as a Collection of Work Sample Tests Change the Problem?**

Lance's suggestion to change the design of ACs toward task- or role-based ACs means treating them as collections of work sample tests. However, it has already been acknowledged in the past that it remains undetermined what work sample tests measure

(e.g., Roth, Bobko, & McFarland, 2005). In their meta-analysis on work sample tests, Roth et al. even said that they "believe that work sample tests could benefit from a continued focus on the constructs they assess" (p. 1031). Thus, turning away from the seemingly troubling construct validity evidence in the AC domain and rethinking ACs as a collection of work samples neither alters the problem nor makes it go away. Instead, important questions remain: What is it that simulation exercises, such as those in ACs or work sample tests, measure? To what degree do these simulations measure the targeted dimensions identified in a previous job analysis? And what leads to successful or less successful completion of these simulations?

Lance ignores the fact that there is already one answer to the question of what leads to successful completion of simulation exercises: Candidates who are able to correctly identify the criteria for success in the different AC exercises (i.e., the targeted dimensions) manage to handle these simulations more successfully and perform better in the AC (Kleinmann, 1993; Preckel & Schüpbach, 2005). Furthermore, the degree to which candidates have correctly identified the targeted dimensions in an AC is predictive (even after controlling for general mental ability) of their performance in a different situation (a selection interview; König, Melchers, Richter, Kleinmann, & Klehe, 2007). Thus, identifying the aspects (i.e., the behavioral dimensions) that are essential in a given situation and directing one's behavior toward

---

Correspondence concerning this article should be addressed to Klaus G. Melchers. E-mail: k.melchers@psychologie.uzh.ch

Address: Psychologisches Institut, Universität Zürich, Binzmühlestrasse 14/12, CH-8050 Zürich, Switzerland  
Klaus G. Melchers and Cornelius J. König, Psychologisches Institut, Universität Zürich

these aspects help a candidate to successfully handle that situation.

Taken together, considering ACs as collections of work samples constitutes only a change of name but does not change the construct validity problem. Furthermore, to focus on tasks and to abandon dimensions would mean taking out the explicit acknowledgment of the criteria that are essential for successful performance in an AC.

### **Is Variance Related to Dimension Factors Irrelevant?**

Lance strongly argues in favor of acknowledging that exercise factors contain true variance. We do agree with this, but we are hesitant to follow the reverse conclusion that dimension variance should be considered irrelevant. This would mean throwing out the baby with the bathwater.

Lance and his collaborators have provided convincing evidence that exercise factors represent true variance and show substantial correlations with external measures (e.g., Lance, Foster, Gentry, & Thoresen, 2004). Similarly, it has been found that assessors' ratings mirror variations in candidates' performance but that construct validity evidence was only established when a candidate's performance profile was consistent across situations but differentiated between dimensions (e.g., Lievens, 2002). These findings are important because they show that assessor ratings are much more accurate than previously thought and capture meaningful variations in candidates' performance and not just method bias.

However, substantial amounts of variance in Lievens' (2001, 2002) studies, for example, were not related to true performance variations. Thus, although assessors' ratings were more veridical than previously thought, they were far from perfect, meaning that it remains important to identify factors that influence the accuracy of such ratings.

Lance does not mention either that Lievens' (2001, 2002) studies only considered very prototypical AC candidates. More specifically, Lievens did not assess whether assessors' PEDRs would capture existing vari-

ations due to exercises as well as variations due to dimensions when both sources of variation are present for a candidate but when the latter are smaller than the former. Therefore, Lievens' studies cannot be taken as evidence that dimension variance is trivial in typical ACs and that dimensions can generally be ignored.

Finally, at least one recent large-scale evaluation of AC construct validity has found evidence that PEDRs reflect dimension factors in addition to exercise factors (Bowler & Woehr, 2006). Even though PEDRs showed higher loadings on exercise factors, dimension factors also accounted for substantial variance. Accordingly, Bowler and Woehr clearly argued against dismissing dimensions.

In summary, although a large amount of relevant variance in PEDRs can be attributed to exercise factors, definitive evidence that PEDRs do not contain any meaningful dimension variance has not yet been presented. Thus, for the time being, it seems premature to suggest that dimension factors can be dismissed.

### **Can Possible Improvements in Construct Measurement in ACs Really Be Ignored?**

Lance concludes that none of the attempts from the past quarter century to improve AC design (i.e., the "design fixes") have come up with strong results. Implicitly, he suggests abandoning this line of research. We disagree because an important question regarding the effects of the suggested design fixes has not yet been answered.

According to the unitarian conceptualization of validity, improvements in construct validity should lead to improvements in criterion-related validity. Even if the effects of design fixes on construct validity are low, these fixes might still result in meaningful improvements in criterion-related validity. Unfortunately, only a single primary study has investigated this assumption for ACs so far: A study by Schleicher, Day, Mayes, and Riggio (2002) found that so-called frame-of-reference rater training for assessors led to improvements in construct validity and to

improvements in criterion-related validity. Further evidence in line with the unitarian conceptualization of validity can be indirectly inferred from two meta-analyses: Psychologists who serve as assessors provide more construct-valid ratings than managers according to Woehr and Arthur's (2003) meta-analysis and more criterion-valid ratings according to Gaugler, Rosenthal, Thornton, and Bentson's (1987) meta-analysis.

Thus, some promising results for design fixes can be found in the literature, and more primary research regarding the joint effects of design fixes on construct and criterion validity is clearly needed—not less—to ensure that we do not miss an important chance to improve ACs.

## References

- Bowler, M. C., & Woehr, D. J. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on assessment center ratings. *Journal of Applied Psychology, 91*, 1114–1124.
- Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72*, 493–511.
- Kleinmann, M. (1993). Are rating dimensions in assessment centers transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology, 78*, 988–993.
- König, C. J., Melchers, K. G., Richter, G. M., Kleinmann, M., & Klehe, U.-C. (2007). The ability to identify criteria in nontransparent selection procedures: Evidence from an assessment center and a structured interview. *International Journal of Selection and Assessment, 15*, 283–292.
- Lance, C. E., Foster, M. R., Gentry, W. A., & Thoresen, J. D. (2004). Assessor cognitive processes in an operational assessment center. *Journal of Applied Psychology, 89*, 22–35.
- 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.
- Lievens, F. (2001). Assessors and use of assessment centre dimensions: A fresh look at a troubling issue. *Journal of Organizational Behavior, 22*, 203–221.
- Lievens, F. (2002). Trying to understand the different pieces of the construct validity puzzle of assessment centers: An examination of assessor and assessee effects. *Journal of Applied Psychology, 87*, 675–686.
- Preckel, D., & Schüpbach, H. (2005). Zusammenhänge zwischen rezeptiver Selbstdarstellungskompetenz und Leistung im Assessment Center [Correlations between receptive self-presentation competence and performance in an assessment center]. *Zeitschrift für Personalpsychologie, 4*, 151–158.
- Roth, P. L., Bobko, P., & McFarland, L. A. (2005). A meta-analysis of work sample test validity: Updating and integrating some classic literature. *Personnel Psychology, 58*, 1009–1037.
- Schleicher, D. J., Day, D. V., Mayes, B. T., & Riggio, R. E. (2002). A new frame for frame-of-reference training: Enhancing the construct validity of assessment centers. *Journal of Applied Psychology, 87*, 735–746.
- Woehr, D. J., & Arthur, W. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management, 29*, 231–258.