

# No Steps Forward, Two Steps Back: The Fallacy of Trying to “Eradicate” Adverse Impact?

WINFRED ARTHUR JR.  
*Texas A&M University*

DAVID WOEHHR  
*University of North Carolina Charlotte*

Within the context of the 50th anniversary of the Civil Rights Act of 1964, and specifically as it pertains to the tenets of Title VII, Lindsey, King, Dunleavy, McCausland, and Jones (2013) state: “This focal article raises and addresses critical issues regarding a yet unanswered question: How can organizational researchers and practitioners contribute to the ultimate goal of eradicating employment discrimination” (p. 391). We argue that in the context of employment testing and selection, at least as per the disparate impact theory of discrimination, this question is the wrong one—certainly as framed by Lindsey et al. To the contrary, instead of holding up the “eradication of employment discrimination” as our ultimate goal, perhaps we should continue to focus on the development, implementation, and support of the best (i.e., most job-related and valid) employment practices possible. This is because employment discrimination, as per the disparate impact theory of discrimination, pertains to the use of employment devices that in the presence of adverse impact, are *not* job related. Moreover, we believe Lindsey et al.’s treatment of this issue suffers from a lack of scientific and technical precision that clouds

the import of their recommendations. This lack of precision is reflected in a number of issues including (a) confusing the distinction between subgroup differences and discrimination, and subgroup differences and adverse impact; and (b) providing a representation of the extant literature that is at odds with our reading of said literature. We address each of these issues in more detail below. It is important to note that although Lindsey et al. address the issue of employment discrimination in several spheres, our comment is primarily focused on their recommendations for the “eradication” of adverse or disparate impact in employment testing and selection.

## **Confusing the Distinction Between Subgroup Differences and Discrimination, and Subgroup Differences and Adverse Impact**

Industrial–organizational (I–O) psychologists have played and continue to play a unique and important role in assisting organizations develop valid selection and other employment-related tools and systems. In fact, a fair amount of recent research highlights the positive impact of high performance work practices on overall organizational performance (Beltrán-Martín, Roca-Puig, Escrig-Tena, & Boullusar, 2008; Crook, Todd, Combs, Woehr, & Ketchen, 2011). However, related to Lindsey et al.’s query and their subsequent

---

Correspondence concerning this article should be addressed to Winfred Arthur, Jr.  
E-mail: w-arthur@tamu.edu

Address: Department of Psychology, Texas A&M University, 4235 TAMU, College Station, TX 77843-4235

recommendations, we think it is important to define some basic concepts and issues that help frame and clarify their recommendations. First, what is “discrimination” within the adverse or disparate impact theory of discrimination? CRA, 1991, Section 105 states that “An unlawful employment practice based on disparate impact is established under this title *only if* (a) a complaining party demonstrates that a respondent uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin *and* the respondent fails to demonstrate the challenged practice is job related for the position in question and consistent with business necessity; or (b) the complaining party makes the demonstration described in subparagraph (C) with respect to an alternative employment practice and the respondent refuses to adopt such alternative employment practice.” (italics added).

And concerning “subparagraph (C)” referred to above, Section 60-3.3B (“Consideration of suitable alternative selection procedures”) of the *Uniform Guidelines* (Equal Employment Opportunity Commission [EEOC], 1978) states that “Where two or more selection procedures are available which serve the user’s legitimate interest in efficient and trustworthy workmanship, and which are *substantially equally valid* for a given purpose, the user should use the procedure which has been demonstrated to have lesser adverse impact” (italics added).

We acknowledge that Lindsey et al. make mention of the above (see p. 400). However, their limited treatment of this and subsequent extensive discussion of “strategies to reduce subgroup differences” belies the import of these two critical boundary conditions and largely confounds the distinction between subgroup differences and discrimination. Consequently, it is worth emphasizing again that the mere presence of subgroup differences is *not* discriminatory. Furthermore, the mere presence of adverse or disparate impact is also not discriminatory. Instead, it is discriminatory if, and only if, it cannot be justified or defended on the

basis of the permissible defenses as outlined in Title VII and the *Guidelines*.

Thus, although it is true that reducing subgroup differences will reduce the likelihood of adverse impact, “eradicating” subgroup differences may also substantially reduce validity. Specifically, if there are construct-relevant differences on the construct of interest (e.g., technical aptitude [Schmidt, 2011]; muscular strength and endurance [The Cooper Institute, 2011]; integrity [Berry, Sackett, & Wiemann, 2007]), then they cannot be removed or eradicated through test design and alteration. To the contrary, better (more valid) measures will capture and reflect these differences more effectively than poorer (less valid) measures. “Subgroup differences are psychological, scientific phenomena that are represented and conceptualized as standardized differences between groups on measures of specified constructs. . . . Adverse impact on the other hand, is a legal and administrative concept, which follows from the logic of the phenomenon of subgroup differences, but is also concerned with the equality of outcomes in real-world decision making” (Arthur, Doverspike, Barrett, & Miguel, 2013, p. 3). Hence, subgroup difference reduction techniques are pretest administration techniques. They are implemented as part of the test design and development process and are predicated not on removing differences between groups on the focal construct (e.g., there is ample physiological theory and empirical evidence for the expected and observed sex-based differences in upper-body strength) but instead on removing observed differences in the focal construct that may arise from construct-irrelevant variance such that, at the end of the day, one can state that the observed differences are real and are not due to an irrelevant construct (that the groups differ on) that is present in the observed scores. Hence, subgroup difference reduction techniques are nothing more than standard good test design and development practices (Arthur et al., 2013). In contrast, adverse impact reduction techniques are posttest administration

techniques and primarily entail attempts to eliminate the differences in observed outcomes after the test has been administered.

### **Misrepresenting the Extant Literature**

In identifying potential strategies for “eradicating” adverse impact in selection, Lindsey et al. make a number of recommendations and summary statements that are at odds with the extant literature. Some of these are noted and very briefly discussed.

*Other selection methods can be used to reduce subgroup differences and, thus, adverse impact.* There is very limited, if any, research that shows that different methods (e.g., interviews, assessment centers, or situational judgment tests) of measuring the same construct (e.g., cognitive ability) result in differences in selection outcomes as per reductions in subgroup differences or adverse impact (cf. Arthur, Edwards, and Barrett [2002] and Schmitt and Mills [2001] who compare different modes of the same method). We do not have *method-based* theories of subgroup differences and, subsequently, adverse impact reduction. Nor is there any empirical evidence (of which we are aware) that supports such a conceptualization, namely a study or studies that have held the construct constant (e.g., general mental ability) and varied the method of measurement, specifically interviews, assessment centers, or situational judgment tests to investigate reductions in subgroup differences with “substantially equal” levels of validity (Arthur & Villado, 2008; Arthur et al., 2013; Bobko & Roth, 2013). To the contrary, it would seem that the “alternative method” approach to adverse impact reduction is really a camouflaged construct-change approach because the switch to alternative methods covaries with changes to more noncognitive constructs (Arthur et al., 2013).

*Banding can be an effective way of reducing subgroup differences.* Lindsey et al. suggest that banding may be an effective

approach to the “reduction of subgroup differences.” Unfortunately, however, the consensus in the extant literature is that the effectiveness of banding as an adverse impact reduction technique is dependent on basing the selection out of bands on the protected class status variable of interest (Barrett, Doverspike, & Arthur, 1995; Bobko & Roth, 2004), a practice that is in violation of Section 106 of CRA 1991. Furthermore, contrary to Lindsey et al.’s statement, banding is an adverse impact reduction technique not a subgroup difference reduction technique (Arthur et al., 2013). Thus, while acknowledging that the “largest reductions in adverse impact are found when subgroup preferences are used within bands” (p. 400), to also state that “research indicates that banding can be an effective way of reducing subgroup differences” (p. 400) is not only incorrect, but it also again confuses the distinction between subgroup differences and adverse impact.

*Alternatives to traditional rank-order decision making.* Similar to banding, Lindsey et al. suggest that there may be other viable alternative approaches to traditional top-down selection that may serve to reduce adverse impact (e.g., using expert based cut scores or differentially weighting predictors based on subgroup impact) without reducing validity. Here it is important to note that, assuming a valid predictor or set of predictors, any deviation from top-down selection will result in a reduction of selection utility (Cascio, Outtz, Zedeck, & Goldstein, 1991).

*Identify and remove items that appear to be biased.* Another potential discrimination reduction strategy recommended by Lindsey et al. “is to identify and remove items that appear to be biased against any marginalized group” (p. 400). Again, this recommendation seriously confounds actual subgroup differences and “bias.” In fact, the limitations and problems inherent in this approach have long been recognized (e.g., Angoff, 1982; Linn & Drasgow, 1987; Lord, 1977). Specifically, whenever two groups are not equal on the construct or trait being assessed, highly “discriminating”

items (in the positive psychometric sense) will appear to be “biased” because they do a better job of differentiating between low-scoring and high-scoring groups. The efficacy of Lindsey et al.’s recommendation is also sample specific; an item that is identified as “biased” in one test administration is not necessarily going to be tagged as “biased” in another. Finally, given the increasingly large number of subgroups and intersectional discrimination (see Arthur et al., 2013), the practical implementation of this approach is inherently challenging and could result in the elimination of a large proportion of the test items. Thus, despite the “intuitive” appeal of this strategy, it represents a seriously flawed approach to the reduction of test bias.

*Increasing minority representation in the applicant pool.* Ironically, contrary to what is suggested by Lindsey et al.’s framing of this issue, simply increasing the number of minority applicants is likely to translate into higher, not lower levels, of adverse impact. As noted by Arthur et al. (2013), “For the same number of minority candidates passing the test, the adverse impact ratio is a function of the number of minority test takers, such that all things being equal, although it is at odds with the typical recruitment practices and goals, a smaller number of minorities will translate into smaller levels of adverse impact” (p. 7). In recognition of this potential unintended effect, Newman and Lyon (2009) discuss the importance and criticality of targeted recruiting (for cognitive ability and conscientiousness) instead of undifferentiated general recruitment. Interestingly, Lindsey et al. make no mention of this critical issue.

## Summary and Conclusion

As a scientific discipline and profession, we need to be realistic about what we can and cannot do and that which is informed by the current extant literature and that which is not. Lindsey et al.’s argument that organizational researchers should focus on “eradicating” employment discrimination is overly broad and vague. If by “eradication

of employment discrimination” we mean the elimination of employment practices that are not job related, then this is a very attainable goal to which I–O psychologists continue to make substantial contributions and progress. However, if by this we mean the elimination of all subgroup differences and adverse impact as is implied by Lindsey et al., then this might best be described as the proverbial quest for the Holy Grail (Arthur et al., 2013; see also McDaniel, Kepes, & Banks, 2011). Specifically, in making attributions about the reasons for observed subgroup differences on employment tests, we need to be clear about whether the observed subgroup differences are due to or caused by the test itself (the source-of-fire hypothesis) or whether the test is merely the indicator rather than the cause of the observed differences (the thermometer hypothesis; see Arthur et al., 2013). It would seem the extant individual and subgroup differences literature is more in accord with the latter perspective rather than the former. So, in the quest to eradicate employment discrimination, consonant with the intent of Title VII, we need to focus on that which we can guarantee as a field and science. We can and should strive for valid selection tests and other employment-related systems and tools—that we *can* do; but much as we wish we could, we cannot guarantee equal outcomes. In summary, we think ultimately, the answer to the question raised by Lindsey et al. is that our goal should be to develop, implement, and support the best (i.e., most job-related and valid) employment practices possible. That *is* what we have control over and can guarantee; we cannot, unfortunately, guarantee the equality of outcomes.

## References

- Angoff, W. H. (1982). Use of difficulty and discrimination indices for detecting item bias. In R. A. Berk (Ed.), *Handbook of methods for detecting test bias* (pp. 96–116). Baltimore, MD: Johns Hopkins University Press.
- Arthur, W., Jr., Doverspike, D., Barrett, G. V., & Miguel, R. (2013). Chasing the title VII Holy Grail: The pitfalls of guaranteeing adverse impact elimination. *Journal of Business and Psychology*. doi: 10.1007/s10869-013-9289-6

- Arthur, W., Jr., Edwards, B. D., & Barrett, G. V. (2002). Multiple-choice and constructed-response tests of ability: Race-based subgroup performance differences on alternative paper-and-pencil test formats. *Personnel Psychology, 55*, 985–1008.
- Arthur, W., Jr., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology, 93*, 435–442.
- Barrett, G. V., Doverspike, D., & Arthur, W., Jr. (1995). The current status of the judicial review of banding: A clarification. *The Industrial-Organizational Psychologist, 33*(1), 39–41.
- Beltrán-Martín, I., Roca-Puig, V., Escrig-Tena, A., & Bou-Llugar, J. C. (2008). Human resource flexibility as a mediating variable between high performance work systems and performance. *Journal of Management, 34*, 1009–1044.
- Berry, C. M., Sackett, P. R., & Wiemann, S. (2007). A review of recent developments in integrity test research. *Personnel Psychology, 60*, 271–301.
- Bobko, P., & Roth, P. L. (2004). Personnel selection with top-score-references banding: On the inappropriateness of current procedures. *International Journal of Selection and Assessment, 12*, 291–298.
- Bobko, P., & Roth, P. L. (2013). Reviewing, categorizing, and analyzing the literature on Black-White mean differences for predictors of job performance: Verifying some perceptions and updating/correcting others. *Personnel Psychology, 66*, 91–126.
- Cascio, W. F., Outtz, J., Zedeck, S., & Goldstein, I. L. (1991). Statistical implications of six methods of test score use in personnel selection. *Human Performance, 4*, 233–264.
- Civil Rights Act of 1964, Public Law 88–352 (78 Stat. 241).
- Civil Rights Act of 1991, 42 U.S.C. CC 1981, 200e *et seq.*
- Crook, T. R., Todd, S. Y., Combs, J. G., Woehr, D. J., & Ketchen, D. (2011). To what extent does human capital shape performance? A meta-analytic examination. *Journal of Applied Psychology, 96*, 443–456.
- Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). Adoption by four agencies of uniform guidelines on employee selection procedures. *Federal Register, 43*, 38290–38315.
- Lindsey, A., King, E., McCausland, T., Jones, K., & Dunleavy, E. (2013). What we know and don't: Eradicating employment discrimination 50 years after the Civil Rights Act. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 6*(4), 391–413.
- Linn, R. L., & Drasgow, F. (1987). Implications of the golden rule settlement for test construction. *Educational Measurement: Issues and Practice, 6*, 13–17.
- Lord, F. M. (1977). A study of item bias using item characteristic curve theory. In Y. H. Poortinga (Ed.), *Basic problems in cross-cultural psychology* (pp. 19–29). Amsterdam, the Netherlands: Swets & Zeitlinger.
- McDaniel, M. A., Kepes, S., & Banks, G. C. (2011). The uniform guidelines are a detriment to the field of personnel selection. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 4*, 494–514.
- Newman, D. A., & Lyon, J. S. (2009). Recruitment efforts to reduce adverse impact: Targeted recruiting for personality, cognitive ability, and diversity. *Journal of Applied Psychology, 94*, 298–317.
- Schmidt, F. L. (2011). A theory of sex differences in technical aptitude and some supporting evidence. *Perspectives on Psychological Science, 6*, 560–573.
- Schmitt, N., & Mills, A. E. (2001). Traditional tests and job simulations: Minority and majority performance and test validities. *Journal of Applied Psychology, 86*, 451–458.
- The Cooper Institute. (2011). *The Cooper Standards*. Dallas, TX: The Cooper Institute. Retrieved from <http://www.cooperinstitute.org/law-fire-military>