Thurstone Might Have Been Right About Attitudes, but Drasgow, Chernyshenko, and Stark Fail to Make the Case for Personality

STEVEN PAUL REISE
University of California, Los Angeles

Drasgow, Chernyshenko, and Stark's (2010) polemic argues the need for, and superiority of, an ideal point response process and an ideal point measurement model, for ''personality'' measurement, broadly defined. Yet, the target article leaves me unconvinced of their merits or necessity in personality trait measurement. The reasons for my skepticism are their (a) questionable distinction between cognitive and noncognitive data as it relates to the applicability and interpretability of dominance response models, (b) weak conceptual link between attitude measurement theory and personality trait assessment, (c) reliance on the findings produced from a single fit index when applied to a limited set of self-report measures, and (d) failure to provide empirical evidence that personality trait scales, in general, fail to include items that provide measurement precision in the ''middle'' or ''intermediate'' level of the trait range.

Cognitive and Noncognitive Measures: Different, but Not That Different

Throughout the target article, a line in the sand is drawn between personality (typical performance) and cognitive (maximal performance) measures and resulting item response data. This distinction is necessary for their arguments because an ideal point response process is hard to fathom for cognitive abilities. Thus, the authors confine their arguments to self-report measures of ''personality'' broadly conceived. Repeatedly throughout the article, personality data and attempts to fit dominance models (e.g., factor analytic or item response theory) to them are referred to in derogatory terms, relative to the apparent shining armor of cognitive tests. For example, ''We believe that dominance models are most sensibly applied to domains in which an individual's capacity or maximum performance capability is pitted against the difficulty or extremity of the item'' (p. 467) and ''. . . psychometric models for dominance response processes, such as classical test theory, factor analysis, and logistic item response theory models, are ill suited for response processes requiring introspection'' (p. 467).

I sympathize with some of their arguments. There are important applied and conceptual differences between cognitive and noncognitive (personality, psychopathology, and patient reported outcomes) measures and resulting item response data (Reise, in press; Reise & Waller, 2003, 2009). For example, many personality and psychopathology constructs are more like unipolar ''quasi-traits,'' definable at only one end of the continuum

Correspondence concerning this article should be addressed to Steven P. Reise.
E-mail: reise@psych.ucla.edu
  Address: Department of Psychology, Franz Hall, UCLA, Los Angeles, CA 90095.

(e.g., depression). In turn, my position is that these differences need serious attention when considering the application of item response theory (IRT) measurement models. Moreover, I agree with the authors that the field of self-report personality assessment, relative to the large-scale cognitive assessment, is burdened by the weight of unexamined trait constructs (Tellegen, 1988, 1991), poorly constructed and psychometrically analyzed scales, score distorting or invalidating response artifacts including method effects, and research that continues with little regard for understanding the underlying response process or the meaning of the latent trait scale (Block, 2010; Borsboom, Mellenbergh, & van Heerden, 2003; Reise & Waller, 2009).

Nevertheless, Drasgow et al. take the cognitive versus noncognitive measurement distinction way too far. First, cognitive tests, such as an eighth grade social studies statewide achievement test, certainly involve an "introspection" component in the form of two fundamental cognitive processes—recall and recognition. Second, and more importantly, of all the well-acknowledged problems in "personality" measurement, there is no compelling evidence that either personality scales in general do not measure well in the "intermediate" range of the construct or that dominance response process models are inadequate. Sweeping statements, such as those cited above, cavalierly dismiss all the exploratory and confirmatory factor analytic research and applications of IRT models to measures of noncognitive constructs that have appeared in well-respected journals (as reviewed in Reise & Waller, 2009). It also ignores the fact that some of psychology's most heavily researched and most validated scales are personality or psychopathology measures developed under a dominance response framework.

Can Personality and Attitude Measurement Be Lumped Together?

The article begins with a famous quote "Attitudes can be measured," yet the authors spend the remainder of the article trying to expand their arguments from attitude measurement to personality measurement more generally. This isn't surprising. Attitude measurement is a small and highly specialized field of limited interest to a broad audience, especially an applied audience; few individuals are hired, fired, or promoted because of their attitudes toward partial birth abortion, gun control, or the length of sentencing for property crimes. Hence, to justify a more general applicability of the ideal point response process, and the associated psychometric model, the authors link personality measurement with attitude measurement. They accomplish this by arguing "personality is an attitude about the self" because personality self-report items require "introspection."

I cannot think of a single personality theorist who would agree with this theory. Consider the items of the California Q-sort (Block, 1961), which taken as a whole is a pretty good definition of the personality domain. Is a tendency toward undercontrol an attitude about the self? How about a high energy level or difficulties handling stress? Moreover, the notion that personality is *merely* "an attitude about the self" does not appear consistent with the empirical evidence that (a) personality measurements of the same construct, either through self-reports, other reports, or behavior tasks, tend to be highly correlated (Furr, Wagerman, & Funder, 2010; Nave, Sherman, & Funder, 2008); (b) personality measures have proven heritabilities (Stubbe, Poshuma, Boomsma, & De Geus, 2005); (c) personality measures are related systematically to biological parameters, such as those assessed via fMRI (Eisenberger & Lieberman, 2004); (d) personality traits are resistant to change and consistent over long periods of time (Friedman et al., 1995); and, most importantly, (e) personality measures correlate meaningfully with important life outcomes (Ozer & Benet-Martinez, 2006).

These empirical facts are more consistent with a view of personality as a stable, biologically based causative mechanism

as opposed to reflecting "attitudes about the self." For example, "We can begin by defining a *trait* as an inferred relatively enduring organismic (psychological, psychobiological) *structure* underlying an extended family of behavioral dispositions. In the case of personality traits it is expected that the manifestations of these dispositions can substantially affect a person's life" (Tellegen, 1991, p. 13). In fact, if one adopts a classic Allportian view of personality traits as being determinative (i.e., influencing a person's perception of a situation), it appears that personality operates to influence a person's introspection processes and thus their attitudes about the self.

If the authors intended to limit their arguments to the cognitive process of responding to self-report personality items, rather than to argue that attitudes and personality traits are essentially the same, I would be more sympathetic to their perspective. Yet I find it puzzling that not a single focus group, case study, or "talk out loud while responding to items" research study was cited that might lend some empirical support that, at least for some self-report personality items, the dominance response does not hold. It seems to me that such fundamental research is necessary and would make an important contribution to personality assessment theory.

Are Ideal Point Models Empirically Supported?

Drasgow et al. argue that their research demonstrates that personality data do not fit dominance IRT models well. Moreover, they argue that their research demonstrates that personality measures fail to include items of "intermediate" difficulty (i.e., provide discrimination in the middle of the trait range), and when such items are written and then fit to ideal point models, such items reveal discrimination ability. In short, the so-called "intermediate" items are not necessarily bad, it is just that psychometricians haven't been fitting the "right" or "superior" model.

Yet the "evidence" to support the need for ideal point models is based on the use of a single chi-square badness-of-fit test and their own rules of thumb for defining "excellent," "satisfactory," and "misfit." More specifically, their studies on the sixteen personality factor and another self-created measure showed that item pairs and triplets displayed a relatively poor fit (i.e., high ratio of chi-square to degrees of freedom). It is not at all clear whether their index provides any meaningful gauge of how wrong the estimated IRT item parameters were or what the applied consequence of such misfit was. It is also not clear whether the misfit was attributable to poor scale construction. For example, personality scales are notorious for containing repeated item content, which, in turn, causes violations of local independence. Such local dependence would almost certainly cause a large chi-square for item triplets or pairs.

It is a leap of faith to argue that badness-of-fit, as judged by a chi-square test, indicates an ideal response process; it may just as well have been generated by multidimensionality, or simply reflect poor item construction. How those competing explanations are evaluated is not clear in the target article. At least dominance response models propose a common latent variable that is "causal" of item variance and thus inter-item covariance (Bollen & Lennox, 1991). Moreover, dominance models allow researchers to make informed judgments as to which items belong to which constructs and which items are "bad." Making such determinations is a complicated and time-consuming process if done seriously (Tellegen & Waller, 2008). I am not convinced that ideal point models afford such empirically based judgments. The examples of items provided, "I enjoy chatting quietly with a friend at a café" (Extraversion) and "My life has had about an equal share of ups and downs" (well-being) leave me questioning whether the proposed underlying latent variables can sensibly explain item responses.

Finally, in support of future measures based on ideal point models, the

authors claim "Some constructs in organizational psychology might be better studied by embracing an ideal point perspective" (p. 472). They then mention person–organization fit, employee performance, organizational commitment, leader behavior, and perceived organizational support. I question whether any latent trait model, including ideal point models, can be usefully applied to these types of socially constructed emergent variables (Bollen & Lennox, 1991). On the other hand, I found the suggestion of reviving forced-choice personality questionnaires at least intriguing, and I am anxious to see if such measures can add anything beyond existing competing measures such as the California Q-set (Block, 1961).

Conclusion

Although I have disputed many of their iconoclastic claims, I agree with the authors on two critical issues. First, it is important to get the psychometric model right, and research should continue that considers the question of what is the right model for specific types of personality measures. Second, we apparently agree that it is critically important to understand the item response process and the nature of the latent variable in personality measures. How are researchers to understand change when they do not understand the processes underlying the items? To the extent that the Drasgow et al. article motivates research along these lines, it has served a useful purpose. To the extent that it motivates researchers to create more redundant and unnecessary self-report "personality" scales, it is a distraction.

References

Block, J. (1961). *The Q-sort method in personality assessment and psychiatric research.* Springfield, IL: Charles C. Thomas.

Block, J. (2010). The five factor framing of personality and beyond: Some ruminations. *Psychological Inquiry, 21*, 2–25.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*, 305–315.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*, 203–219.

Drasgow, F., Chernyshenko, O. S., and Stark, S. (2010). 75 years after Likert: Thurstone was right! *Industrial and Organizational Psychology: Perspectives on Science and Practice, 3*, 465–476.

Eisenberger, N. I., & Lieberman, M. D. (2004). Why rejection hurts: A common neural alarm system for physical and social pain. *Trends in Cognitive Sciences, 8*, 294–300.

Friedman, H. S., Tucker, J. S., Schwartz, J. E., Tomlinson-Keasey, C., Martin, L. R., Wingard, D. L., & Criqui, M. H. (1995). Psychosocial and behavioral predictors of longevity: The aging and death of the "Termites." *American Psychologist, 50*, 69–78.

Furr, R. M., Wagerman, S. A., & Funder, D. C. (2010). Personality as manifest in behavior: Direct behavioral observation using the revised Riverside Behavioral Q-Sort (RBQ-3.0). In C. R. Agnew, D. E. Carlston, W. G. Graziano, & J. R. Kelly (Eds.), *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (pp. 186–204). New York: Oxford University Press.

Nave, C. S., Sherman, R. A., & Funder, D. C. (2008). Beyond self-report in the study of hedonic and eudaimonic well-being: Correlations with acquaintance reports, clinician judgments and directly observed social behavior. *Journal of Research in Personality, 42*, 643–659.

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 52*, 401–421.

Reise, S. P. (in press). The emergence of item response theory (IRT) models and the Patient Reported Outcomes Measurement Information System (PROMIS). *Austrian Journal of Statistics.*

Reise, S. P., & Waller, N. G. (2003). How many IRT parameters does it take to model psychopathology items. *Psychological Methods, 8*, 164–184.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology, 5*, 27–48.

Stubbe, J. H., Poshuma, D., Boomsma, D. I., & De Geus, J. C. (2005). Heritability of life satisfaction in adults: A twin-family study. *Psychological Medicine, 35*, 1581–1588.

Tellegen, A. (1988). The analysis of consistency in personality assessment. *Journal of Personality, 56*, 621–663.

Tellegen, A. (1991). Personality traits: Issues of definition, evidence and assessment. In D. Cichetti, & W. Grove (Eds.), *Thinking clearly about psychology: Essays in honor of Paul Everett Meehl* (pp. 10–35). Minneapolis: University of Minnesota Press.

Tellegen, A., & Waller, N. G. (2008). Exploring personality through test construction: Development of the multidimensional personality questionnaire. *The Sage handbook of personality theory and assessment.*In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *Handbook of personality theory and testing: Vol. II. Personality measurement and assessment* (pp. 254–285). London: Sage.