Industrial and Organizational Psychology, **3** (2010), 498–501.
Copyright © 2010 Society for Industrial and Organizational Psychology. 1754-9426/10

Six Questions That Practitioners (Might) Have About Ideal Point Response Process Items

DEV K. DALAL AND SCOTT WITHROW
Bowling Green State University

ROBERT E. GIBBY
Procter & Gamble

MICHAEL J. ZICKAR
Bowling Green State University

The ideal point model discussed in Drasgow, Chernyshenko, and Stark (2010) has the potential to advance the understanding of item responses, and the method has significant implications for psychometric practice. Although Drasgow et al. outlined some avenues for applying the results of the ideal point response process, questions and challenges about applying these models in real-world psychometric settings remain. The goal of this commentary is to present questions that industrial–organizational (I–O) practitioners may have about using ideal point measurement in their work. These six questions highlight gaps in our current knowledge that will require further research to adequately address.

Correspondence concerning this article should be addressed to Dev K. Dalal.
E-mail: ddalal@bgsu.edu
Address: Department of Psychology, Bowling Green State University, Bowling Green, OH 43403.
Dev K. Dalal and Scott Withrow, Department of Psychology, Bowling Green State University; Robert E. Gibby, Procter & Gamble; Michael J. Zickar, Department of Psychology, Bowling Green State University.

Question #1: How Can an Ideal Point Measure Be Scored Without IRT-Based Parameters?

The original impetus for Likert (1932) to develop his scoring method was the difficulty associated with scoring Thurstone's (1928) original measures. Thurstone's method also required a large calibration sample to scale the items on the latent trait distribution. Likert's method of scoring was simpler to create and use (e.g., simply adding scores) and did not require a large sample size. Later, advances in statistical techniques like Roberts and Laughlin's (1996) development of an item response theory (IRT) model to analyze ideal point measures helped solve some of the issues facing researchers like Likert and Thurstone. Applying these IRT models to ideal point measures, researchers can estimate respondents' standing on the

https://doi.org/10.1111/j.1754-9434.2010.01279.x Published online by Cambridge University Press

latent trait continuum. However, applying these IRT models still requires large sample sizes, and assessing unidimensionality with these models is difficult and understudied. Drasgow et al.'s recommendation of calculating the mean of item locations also requires a calibration sample that an I–O practitioner may not have. For practitioners operating on a global scale, this presents an even more severe issue as multiple calibration samples would be required to ensure a single scoring algorithm or to develop culturally based scoring algorithms.

Given these limitations to existing scoring methods, the first question we believe a practitioner may have is ''How do we score ideal point measures without the use of calibration samples or IRT?'' That is, if I–O practitioners have an *N* of 100, how can we score an ideal point measure? With traditional Likert scoring systems, more is always better—except for reverse-coded items—with more positive options receiving higher scale values. With ideal point items, the relationship between scale options and scale values may be nonlinear. Rubrics for understanding these scoring schemes will be essential, especially for practitioners who do not have access to, or the capability for, running IRT models to score respondents.

Question #2: What Should Practitioners Do If They Want to Create an Ideal Point Scale, but They Only Have a Small Sample Size?

Beyond requiring large sample sizes to score ideal point measures, practitioners need large sample sizes to develop and evaluate ideal point scales. Applying IRT models to scales requires sample sizes large enough to obtain stable parameter estimates. As the number of item parameters within a model increases, the sample size required to estimate those parameters with reasonable precision increases. A practitioner will need a larger sample size to fit an ideal point IRT model, as opposed to a dominance model. Roberts, Donoghue, and Laughlin (2000) suggest that a minimum of 750 participants are needed to uncover stable parameter estimates for a 20-item measure, making it difficult to apply the generalized graded unfolding model (GGUM) to measures. Even if one chooses to utilize Thurstone's method of scaling items, a sample size of nearly 300 is needed (per Thurstone's [1928] recommendation). Regardless of whether practitioners wish to use IRT modeling or Thurstone's original scaling method, they must obtain a large sample—likely larger than the applicant pool at their disposal—to evaluate an ideal point measure. Similar to scoring ideal point measures with a limited sample size, practitioners would benefit from any research or advice on creating and evaluating ideal point scales when only a limited sample size is available.

Question #3: How Do We Write Ideal Point Items?

Although the increase in precision that ideal point measures provide is positive, guidelines for creating ideal point items have yet to be developed. Indeed, Chernyshenko, Stark, Drasgow, and Roberts (2007) noted that writing ideal point items is difficult because (a) no classical test theory tools exist to evaluate quality of items and (b) Thurstone's original scaling method (discussed above) was cumbersome (p. 92).

Guidelines need to be written to provide item writers a taxonomy of discriminating qualities that produce items with significant unfolding. Many of the item-writing heuristics taught in tests and measures courses are based on the dominance model. For example, the heuristic of avoiding double-barreled items may no longer be appropriate for ideal point items.

In addition to guidelines on how to write good unfolding items, researchers will need to change the current method of evaluating items. Researchers who develop scales with dominance theory assumptions have the

https://doi.org/10.1111/j.1754-9434.2010.01279.x Published online by Cambridge University Press

luxury of using the tools of classical test theory (CTT). CTT provides researchers with guidance on good and bad items in small pilot samples, a practice with no analogue for unfolding models.

Question #4: What Are the Implications for Scale Length Using Ideal Point Assumptions to Create Scales?

Related to writing effective ideal point items, little information exists about the requirements for scale length created with ideal point items to achieve reasonable estimates of standing on the latent trait. Similar issues regarding scale length have been debated in the literature related to dominance items, and some guidelines have been suggested based on CTT. For example, it has been suggested that as few as three items per scale can be used to obtain an adequate level of internal consistency (see Hinkin, 1998 for review), or a minimum of three to four items are needed to adequately scale a latent construct in structural equation modeling (Kline, 2005). Finally, one can estimate the number of (parallel) items needed to obtain a level of reliability by applying the Spearman–Brown formula (Allen & Yen, 1979).

Ideal point measures need to be longer because of the need to untangle responses from people throughout the range of the trait distribution. Respondents who disagree on a series of items may be either extremely low or extremely high on a particular trait; it is only in conjunction with more moderate items that these respondents can be differentiated. Given recent trends in industry for creating shorter tests to address concerns from candidates and business partners (e.g., recruiters), the creation of longer selection instruments or employee surveys could be problematic. Therefore, like advice on writing items, both I–O practitioners and researchers would benefit from greater understanding of how using ideal point assumptions to create scales will affect scale length.

Question #5: In the Noncognitive Area, Would a Practitioner Expect to See Higher Criterion-Related Validity Coefficients Through the Adoption of Ideal Point Measures?

Drasgow et al. mention that the adoption of ideal point approaches to the measurement of noncognitive domains results in better measurement precision of that construct. However, an important question is whether this increase in measurement precision results in stronger relationships between constructs. Given how resource intensive it is to create ideal point measures (e.g., large sample size and more difficult item writing), practitioners should be assured that their investments will yield returns. Therefore, researchers should look to establish that ideal point measure estimates of constructs are more strongly related to criteria than a simple sum score of dominance items. In one example, Chernyshenko et al. (2007) showed that their ideal point measure of orderliness had stronger criterion-related validities than did dominance measures.

When evaluating the relation of ideal point measures to external criteria, one must remember that validity coefficients are sensitive to the whole range of trait distribution. Therefore, it is possible that the correlation may change a little compared to dominance measures, but top-down selection decisions may differ because of the significant changes at the top end of the distribution. Therefore, researchers should not just focus on correlation coefficients when examining how ideal point models influence hiring decisions. I–O research can further inform I–O practice about the changes in validity associated with using ideal point measures to determine whether the benefits of using ideal point measures outweigh the costs associated with creating one.

Question #6: What Is the Applicability of Using the Ideal Point Model to Develop Measures of Engagement and Satisfaction?

Employee satisfaction and engagement are two constructs that are very popular in

https://doi.org/10.1111/j.1754-9434.2010.01279.x Published online by Cambridge University Press

applied and academic I–O work. Often measures of these constructs were developed using dominance models, making it difficult to interpret the midpoint of the scale. In other words, if a respondent is scored as ''neutral'' on an engagement scale, does this mean that this employee is moderately engaged or that she is ambivalent in her engagement? The ideal point response process of comparing one's standing on a latent distribution of the construct to the location of the item on the latent distribution—what Drasgow et al. referred to as introspection—seems to fit the measurement of engagement and satisfaction better than the dominance approach. Preliminary evidence of this response process underlying satisfaction surveys may be obtained from Carter and Dalal (2010), showing that the Job Descriptive Index Work scale contained ideal point items.

If ideal point models measure these constructs better, it is possible that new questions about these constructs might be asked. For example, practitioners would no longer have trouble interpreting ''neutral'' scores. In addition, applying ideal point assumptions to scales of these constructs might yield more precise measurement. With an increase in measurement precision, I–O practitioners gain a better understanding of the employees in organizations and take action to improve their lives.

Conclusions

The study of ideal point response processes could still be considered in its preliminary stages. Although posited by Thurstone (1928) over 80 years ago, Likert's (1932) more straightforward and less labor-intensive approach to measurement dominated. With the resurgence of interest in ideal point response models and new approaches to analyzing these measures (e.g., GGUM), it seems reasonable to suspect that ideal point measures will see favor in the future. How long in the future is a matter of showing researchers and practitioners the value of this method. Indeed, Drasgow et al. lamented that it could possibly take as long as 30 years to see the widespread adoption of ideal point measures. The research on ideal point models is exciting and, as Drasgow et al. show, will likely revolutionize measurement practice and theory. We hope the questions we identified will be addressed by researchers as we move forward in our understanding of the ideal point response process.

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Long Grove, IL: Waveland Press.

Carter, N. T., & Dalal, D. K. (2010). An ideal point account of the JDI Work satisfaction scale. *Personality and Individual Differences, 49*, 743–748.

Chernyshenko, O. S., Stark, S., Drasgow, F., & Roberts, B. W. (2007). Constructing personality scales under the assumptions of an ideal point response process: Toward increasing the flexibility of personality measures. *Psychological Assessment, 19*, 88–106.

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

Hinkin, T. R. (1998). A brief tutorial on the development of measures for us in survey questionnaires. *Organizational Research Methods, 1*, 104–121.

Kline, R. B. (2005). *Principle and practice of structural equation modeling* (2nd ed.). New York: The Guilford Press.

Likert, R. (1932). The method of constructing an attitude scale. *Archives of Psychology, 140*, 44–53.

Roberts, J. S., Donoghue, J. R., & Laughlin, J. E. (2000). A generalized item response theory model for unfolding unidimensional polytomous responses. *Applied Psychological Measurement, 24*, 3–32.

Roberts, J. S., & Laughlin, J. E. (1996). A unidimensional item response model for unfolding responses from a graded disagree–agree response scale. *Applied Psychological Measurement, 20*, 231–255.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 33*, 529–554.

https://doi.org/10.1111/j.1754-9434.2010.01279.x Published online by Cambridge University Press