Industrial and Organizational Psychology, **3** (2010), 477–480.

COMMENTARIES

Questionable Defeats and Discounted Victories for Likert Rating Scales

CHRISTOPHER J. WAPLES, WILLIAM S. WEYHRAUCH, ANGELA R. CONNELL, AND SATORIS S. CULBERTSON
Kansas State University

In their focal article, Drasgow, Chernyshenko, and Stark (2010) depict Thurstone scaling methods as superior to Likert rating scales, particularly for attitude assessment, noting numerous benefits of the former. In their fervor to give credit to Thurstone scaling methods however, they tend to discount the benefits of Likert scaling, leaving the reader to question the utility of Likert scaling in any case. We believe that discarding Likert scales for attitude measurement, a suggestion that, although not explicitly stated, seems to be implied, would be premature and akin to throwing the baby out with the bathwater.

Questionable Victories for Thurstone (and Defeats for Likert)

Drasgow et al. highlight numerous benefits of Thurstone scaling relative to Likert scaling. On the surface, their detailed analysis provides compelling support for the use of ideal point methods for attitude assessment. However, although we concede the superiority of Thurstone scaling in some settings, we suggest that some of the benefits they note are not so one-sided and that, instead, the jury is still out on whether Thurstone or Likert methods come out ahead. Below, we discuss the flaws in their assertions that ideal point models more closely match the introspective decision process required by attitude measurement and that the capacity to accurately measure high, low, *and* intermediate levels of the latent trait is specific to ideal point methods.

Matching Introspective Decision Processes

Foremost among Drasgow et al.'s justifications for reconsidering the Thurstone approach is the suggestion that the ideal point model more closely matches the introspective decision process required by attitude measurement. In this model, an individual's responses ensue from comparing perceptions of item content to relevant self-perceptions, and the resulting decision reflects the distance between the two (Drasgow et al.). Although this explanation has merit, the question remains as to whether this process is truly distinct from Likert scaling, and as such whether there is a clear ''victory'' for Thurstone methods over Likert methods. That is, Drasgow et al. have interpreted Likert's approach within a dominance framework, which suggests that responses are the result of a dominant latent trait. The authors did not, however, expand on the underlying process by which a dominant trait leads to

Correspondence concerning this article should be addressed to Christopher J. Waples.
E-mail: cwaples@ksu.edu
Address: Department of Psychology, Kansas State University, 558 Bluemont Hall, 1100 Mid-Campus Drive, Manhattan, KS 66506–5302
The authors would like to thank Patrick Knight for his comments on an earlier version of this commentary.
Christopher J. Waples, William S. Weyhrauch, Angela R. Connell, and Satoris S. Culbertson, Department of Psychology, Kansas State University.

https://doi.org/10.1111/j.1754-9434.2010.01274.x Published online by Cambridge University Press

such decisions for noncognitive assessment. A comparison of item content perceptions to self-perceptions, not unlike the process described for the Thurstone method, may serve as the mechanism through which the latent trait expresses its dominance.

Even mechanically, the two methods differ primarily as a function of habit rather than of necessity. As Drasgow et al. note, a more linear, cumulative scoring system is typically used for applications of Likert scaling. However, the ''unfolding technique'' encouraged by Drasgow et al. might be used with data derived from Likert scales. Andrich (1996), for example, stated that ''there is nothing inherent in Likert-style questionnaires which makes the response formats cumulative rather than unfolding'' (p. 362). As such, the superiority of Thurstone scaling in matching introspective decision processes remains debatable.

Measuring Intermediate Levels of the Latent Trait

Another questionable victory of Thurstone scaling over Likert scaling lies in its capacity to accurately measure high, low, *and* intermediate levels of the latent trait (Drasgow et al.; Roberts, Laughlin, & Wedell, 1999). In particular, the authors assert that Likert scales imprecisely measure respondents with intermediate trait levels because of the frequent absence of intermediately stated items. We dispute, however, as with the issues regarding matching introspective decision processes, that the use of intermediate items is unique to Thurstone scaling. Although most Likert scales do not include such items, it *is* possible to do so (Andrich, 1996). Whether they *should* be included, or whether there is another way to access intermediate trait levels, is another story. Certainly, their inclusion would reduce traditional scale reliability estimates. Item response theory may present an alternative approach, but given the response options for each item on a Likert-based instrument, scaling on the basis of theta becomes especially convoluted.

Furthermore, the presumption that an absence of intermediate items indicates an inability to access intermediate trait levels seems erroneous. Likert scales typically distinguish between individuals who are low, moderate, or high on a trait by including a range of response options that encompass the trait continuum (Likert, 1932). Ignoring the conventions of Likert scaling, Drasgow et al. have conveniently presented dominance (Likert) items in the form of a dichotomous response format (see Figure 3 in the focal article), thus effectively creating a ''straw man'' for them to nudge over. Just as it is not desirable to arbitrarily dichotomize variables in practice (MacCallum, Zhang, Preacher, & Rucker, 2002), Likert items are not meant to be dichotomously scored. Rather than examining a single figure with the probability of a positive response graphed against theta (necessary for a dichotomous item), a more realistic examination would require comparisons of the probability of endorsement to theta at each point on the anchored response scale provided for a single Likert item.

Imagine that you have a Likert-based measure of Extraversion with five points (strongly disagree to strongly agree). Considering the ''strongly disagree'' option, a highly introverted person (theta $= -3.0$ in Drasgow et al.'s Figure 4) would have a high probability of endorsement. The peak of the response curve would be at the far left, dropping close to zero as theta increased, yielding a strong positive skew. For the ''strongly agree'' option, the opposite would occur. Neither extreme introverts nor extreme extraverts would be likely to endorse the neutral option (''3'' on the scale), but those with a theta near zero should endorse that option more frequently than either extreme option, yielding a curve similar to Drasgow et al.'s Figure 4. With Likert scales having intermediate responses available on every item, it is difficult to see how Drasgow et al.'s assertion that Thurstone scales exclusively provide access to intermediate trait levels remains tenable.

https://doi.org/10.1111/j.1754-9434.2010.01274.x Published online by Cambridge University Press

Finally, Drasgow et al. unequivocally equate Likert scaling with the dominance response model and then proceed to list the problems with dominance models. We feel that this is a confounding of the Likert scaling approach and the dominance response model. Although the authors' main goal may have been to point out the flaws of the dominance model, their manuscript's title and pairing of Likert scales with the dominance response process seem to cement Likert scaling as the underlying focus of their discussion of modeling.

Discounted Victories for Likert (and Defeats for Thurstone)

Although Drasgow et al. briefly acknowledged some ways in which Likert techniques are advantageous to Thurstone techniques, these advantages were largely overlooked by Drasgow et al. in favor of further discussion of Thurstone's approach. In the spirit of fair competition, a more thorough presentation of the benefits of Likert scaling must be offered. Accordingly, we provide further discussion of these benefits, particularly increased reliability and practical utility of Likert techniques over Thurstone techniques.

Increased Reliability

Likert rating scales appear to have an advantage over Thurstone techniques in higher test reliability. Although Drasgow et al. note that Likert cited such evidence in support of his method (Likert, 1932), the point was only mentioned briefly and implied that only Likert himself found such evidence. In reality, Likert's assertion of greater reliability per item has been supported across decades of research. Studies comparing the psychometric properties of Thurstone and Likert scales have typically found reliability estimates in the 0.90s for Likert scales and estimates in the 0.80s for Thurstone scales (Edwards & Kenney, 1946; Poppleton & Pilkington, 1963; Rhoads & Landy, 1973).

Increased Practicality

Perhaps the greatest advantage of Likert scaling is the simplicity and practicality of developing the scale. Drasgow et al. noted the discrepancy, although more in passing than in acknowledgment of its importance. They note that ''Likert provided a much simpler alternative to Thurstone scaling'' and admit to the impracticality of ideal point techniques, stating that ''currently available psychometric models for ideal point data are considerably more complicated than corresponding models for dominance data.'' Although simplicity is not always desirable for its own sake, in the case of attitude assessment it translates to very real pragmatic value.

The increased burden of generating and calibrating items for a Thurstone scale is not insignificant. Edwards and Kenney (1946) determined that development of Thurstone scales requires approximately twice as much time as development of Likert scales, not including the time spent by third-party judges performing Q sorts. Likewise, Barclay and Weaver (1962) found that it took approximately 134 hours to construct a Thurstone scale but just less than 94 to construct a similar Likert scale.

Conclusions

In sum, we believe it is premature to suggest, or imply, that Likert techniques should be abandoned for Thurstone techniques in all cases of attitude measurement. At the very least, the ease of constructing Likert scales, along with their consistently higher estimates of reliability and greater intuitive appeal among organizational decision makers, suggests that they still have a place within industrial and organizational psychology.

References

Andrich, D. (1996). A hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. *British Journal of Mathematical and Statistical Psychology, 49*, 347–365.

https://doi.org/10.1111/j.1754-9434.2010.01274.x Published online by Cambridge University Press

Barclay, J. E., & Weaver, H. B. (1962). Comparative reliabilities and ease of construction of Thurstone and Likert attitude scales. *The Journal of Social Psychology*, *68*, 109–120.

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.

Edwards, A. L., & Kenney, K. C. (1946). A comparison of the Thurstone and Likert techniques of attitude scale construction. *Journal of Applied Psychology*, *30*, 72–83. doi: 10.1037/h0062418

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*, 1–55.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*, 19–40. doi: 10.1037/1082-989x.7.1.19

Poppleton, P. K., & Pilkington, G. W. (1963). A comparison of four methods of scoring an attitude scale in relation to its reliability and validity. *British Journal of Social and Clinical Psychology*, *3*, 36–39.

Rhoads, R. F., & Landy F. J. (1973). Measurement of attitudes of industrial work groups toward psychology and testing. *Journal of Applied Psychology*, *58*, 197–201. doi: 10-1037/h0035653

Roberts, J. S., Laughlin, J. E., & Wedell, D. H. (1999). Validity issues in the Likert and Thurstone approaches to attitude measurement. *Educational and Psychological Measurement*, *59*, 211–233. doi: 10.1177/00131649921969811