

# Response-Scale Formats and Psychological Distances Between Categories

Applied Psychological Measurement  
2016, Vol. 40(1) 73–75  
© The Author(s) 2015  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/0146621615597961  
apm.sagepub.com



Tyler Hamby<sup>1</sup> and Daniel S. Levine<sup>1</sup>

## Keywords

ordinal scales, partial credit model, rating scales

It has been argued that—for Likert-type scales with only the endpoints labeled (END)—respondents assume that the middle, numerically labeled categories are equally spaced and should view the scale as being interval (Schaeffer & Presser, 2003). Conversely, fully labeled (ALL) scales are generally thought to be ordinal. This brief note replicates a recent study on the “psychological distance” between response categories at varying scale lengths (Wakita, Ueshima, & Noguchi, 2012), and it extends the study by considering label format and the impact of response styles.

Wakita et al. (2012) used the generalized partial credit model (GPCM; Muraki, 1992) on ALL scales with four, five, or seven response options ( $m$ ) to estimate  $m$  scale values ( $S_p$ ). These values estimate the perceived “psychological location” of the various alternatives along the response scale (see Appendix A of the online supplementary material for details). They found that the average psychological distance from these converted scale values ( $S_p$ ) to the categories’ numbers ( $p$ ) increased with the total number of categories ( $m$ ) from four to seven.

In this study, 882<sup>1</sup> Amazon Mechanical Turk workers were randomly assigned to either ALL ( $n = 466$ ) or END ( $n = 416$ ) survey conditions. Each respondent was given ten 4-point, ten 5-point, and ten 7-point items; each scale had 5 extraversion and 5 neuroticism items from the International Personality Item Pool (IPIP; 2001). The category labels—1 = *very inaccurate*, 2 = *inaccurate*, 3 = *moderately inaccurate*, 4 = *neither inaccurate nor accurate*, 5 = *moderately accurate*, 6 = *accurate*, and 7 = *very accurate*—were chosen based on research showing that people perceive these labels as equidistant. For 5-point scales, the second and sixth categories were removed; for 4-point scales, the middle category was also removed. For both ALL and END conditions, two GPCM analyses were run—one for the extraversion items and one for the neuroticism items—and converted scale values were computed ( $S_p$ ; see Table 1 for averages). Then, for each item in both conditions (ALL and END), the average difference ( $\bar{D}$ ) between the converted scale values and the category numbers was found.

An itemmetric ANOVA on these psychological distances ( $\bar{D}$ ) was run, meaning that the ( $n = 30$ ) items comprised the sample rather than the participants. Label format (ALL, END) was a within-item variable. The construct measured (extraversion, neuroticism) and number of

<sup>1</sup>The University of Texas at Arlington, USA

## Corresponding Author:

Daniel S. Levine, Department of Psychology, University of Texas at Arlington, 701 S. Nedderman Drive, Arlington, TX 76019-0528, USA.

Email: levine@uta.edu

**Table 1.** Average Converted Scale Values for ALL and END Response Scales With Four, Five, and Seven Response Categories.

Scale format	S <sub>1</sub>	S <sub>2</sub>	S <sub>3</sub>	S <sub>4</sub>	S <sub>5</sub>	S <sub>6</sub>	S <sub>7</sub>
Extraversion							
4-Categories/ALL	1.00	2.01	2.96	4.00			
4-Categories/END	1.00	2.09	3.04	4.00			
5-Categories/ALL	1.00	2.37	2.92	3.57	5.00		
5-Categories/END	1.00	2.14	2.97	3.81	5.00		
7-Categories/ALL	1.00	2.11	3.12	3.68	4.78	5.61	7.00
7-Categories/END	1.00	1.98	2.78	3.63	5.25	6.19	7.00
Neuroticism							
4-Categories/ALL	1.00	2.01	2.94	4.00			
4-Categories/END	1.00	1.96	2.84	4.00			
5-Categories/ALL	1.00	2.37	2.92	3.57	5.00		
5-Categories/END	1.00	2.25	3.05	3.84	5.00		
7-Categories/ALL	1.00	2.77	3.92	4.21	5.26	6.68	7.00
7-Categories/END	1.00	2.34	3.42	4.14	5.52	6.82	7.00

response categories (four, five, seven) were between-item variables. As predicted, the ALL condition ( $M = 0.36$ ,  $SE = 0.03$ ) had significantly greater average  $\bar{D}$  than the END ( $M = 0.27$ ,  $SE = 0.03$ ) condition, Wilks's  $\lambda = 0.77$ ,  $p < .01$ ,  $\eta^2 = .23$ . There was no main effect for construct, but there was a main effect for response-scale length,  $F(2, 26) = 22.05$ ,  $p < .001$ ,  $\eta^2 = .63$ . Using Bonferroni adjustments, 4- ( $M = 0.13$ ,  $SE = 0.04$ ) and 5-point ( $M = 0.29$ ,  $SE = 0.04$ ) scales differed,  $p = .04$ ; 5- and 7-point ( $M = 0.53$ ,  $SE = 0.04$ ) scales differed,  $p < .01$ ; and 4- and 7-point scales differed,  $p < .001$ . However, there was no significant interaction between label format and response-scale length, and there was no interaction between label format and construct.

Net acquiescence (NARS; the tendency to agree with items regardless of content) and extreme response styles (ERS; the tendency to disproportionately use extreme response categories) were considered additionally, as defined by Weijters, Cabooter, and Schillewaert (2010), because these have been shown to be associated with scale formats. Using multilevel linear modeling, the impact that several predictors—NARS, ERS, label format, response-scale length, and the interaction between label format and response-scale length—had on individual differences in psychological distances between individualized scale values and the selected response options (see Appendix B of the online supplementary material for details) was examined. The intraclass correlation ( $ICC = .22$ ) suggests that there were large individual differences in distance. To briefly summarize the results (see Appendix C of the online supplementary material for full results), label format had a minor effect,  $b = -0.09$ ,  $\beta = -.10$ ,  $p < .01$ ; response-scale length had a very strong effect,  $b = 0.40$ ,  $\beta = .54$ ,  $p < .001$ ; there was a weak interaction between response-scale length and label format,  $b = -0.32$ ,  $\beta = -.02$ ,  $p < .05$ ; and ERS had a strong effect,  $b = 0.29$ ,  $\beta = .33$ ,  $p < .001$ . Thus, END scales had slightly smaller distance, but response-scale length and ERS were both strongly, positively associated with distance.

This study supports the conclusion that people perceive endpoint only scales as being interval, at least more than for fully labeled scales, and, replicating Wakita et al. (2012), as the number of categories increases, peoples' perceptions of the response categories become less accurate. Furthermore, in individual differences analyses, ERS was associated with greater psychological distance and, in analyses not shown, with fewer response categories (see also Weijters et al., 2010). Thus, one must find a compromise between potentially increasing the

imprecision associated with ordinal scales by using too many categories and inviting the error of extreme responses by using too few categories. For this purpose, using 5-point END scales is recommended.

### **Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### **Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was partly supported by Grant N000141310562 from the Office of Naval Research to the first author.

### **Note**

1. Eleven cases were screened out for excessive missing values and signs of giving random responses.

### **Supplemental Material**

The online appendix is available at <http://apm.sagepub.com/supplemental>

### **References**

- International Personality Item Pool. (2001). *A scientific collaboratory for the development of advanced measures of personality traits and other individual differences*. Available from <http://ipip.ori.org/>
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*, 159-176.
- Schaeffer, N. C., & Presser, S. (2003). The science of asking questions. *Annual Review of Sociology, 29*, 65-88.
- Wakita, T., Ueshima, N., & Noguchi, H. (2012). Psychological distance between categories in the Likert scale: Comparing different numbers of options. *Educational and Psychological Measurement, 72*, 533-546.
- Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format in response styles: The number of response categories and response category labels. *International Journal of Research in Marketing, 27*, 236-247.