Semantics and Scale: The Impact of Polarity and Wording on Performance Evaluation

Carrie A. Picardi, Ph.D.
Department of Management
Ernest C. Trefz School of Business
University of Bridgeport, Bridgeport, CT

Abstract

This study examined the impact of measurement scale wording on rater judgment and leniency in an employee performance evaluation context. Participants evaluated ratees in a task simulation video using a five-point anchored scale with either unipolar, bipolar, or no anchor labels. Findings partially supported the hypotheses, suggesting scale descriptors may affect performance rating accuracy.

Introduction

• The structure and wording of measurement scales, an important focus in survey research, is also relevant to organizations in the context of employee performance appraisals. Researchers have found that evaluators are often influenced by features such as scale polarity and the words used as anchor labels which may affect their judgment (Barnette, 2000; Tourangeau, Couper, & Conrad, 2007).

• The use of unipolar scales, typically consisting of positively worded anchor labels (e.g., 1 = not satisfied at all, 5 = completely satisfied) or bipolar scales, typically consisting of both positively and negatively worded anchor labels (e.g., 1 = completely dissatisfied, 5 = completely satisfied) may influence rater interpretation and response, and ultimately the validity of the data provided (Horan, DiStefano, & Motl, 2003). Mazaheri and Theuns (2009) investigated the effects of scale polarity and anchoring to examine the impact on participants' responses to a life satisfaction survey. Results demonstrated that ratings were positively skewed for unipolar scales and negatively skewed for bipolar scales for dissatisfaction. This suggests that participants may respond to questions differently depending on if they were presented with a unipolar or a bipolar scale.

• The present study examined the following hypotheses:

Hypothesis 1: Participants in the bipolar-anchor label scale condition will demonstrate a significantly greater degree of performance rating leniency compared to participants in the unipolar-anchor label scale condition.

Hypothesis 2: Participants in the bipolar-anchor label scale condition will demonstrate a significantly greater degree of performance rating leniency compared to participants in the no-anchor label scale condition.

Method

• Participants (n = 128) were randomly assigned to one of three experimental conditions: (1) bipolar-anchor label scale, (2) unipolar-anchor label scale, (3) control no-anchor label scale. They viewed a 10-minute video that presented a group of three individuals (ratees) working through a problem-solving simulation exercise.

• Participants were instructed to evaluate the individuals on behaviors demonstrated across the performance dimensions of collaboration, verbal communication, and decision making and provide rating scores for each ratee in each performance dimension.

• Following the rating score completion, the participants were debriefed and adjourned.

Results

• A series of one-way ANOVAs were conducted to analyze rating scores within the three conditions. Results indicated rating scores in the bipolar-anchor scale condition were significantly higher than rating scores in the unipolar-anchor scale or control conditions for the first ratee in the verbal communication performance dimension, $F(2, 127) = 3.077$, $p<.05$ (Table 1), and in the decision making performance dimension $F(2, 127) = 4.542$, $p<.05$ (Table 2).

• The bipolar-anchor label condition yielded a significantly greater degree of rating leniency than either the unipolar-anchor label or control conditions for one of the three ratees in two of the three performance dimensions, thus partially supporting the hypothesis.

Figure 1: Mean Rating Scores for Ratee 1 in Communication and Decision Making
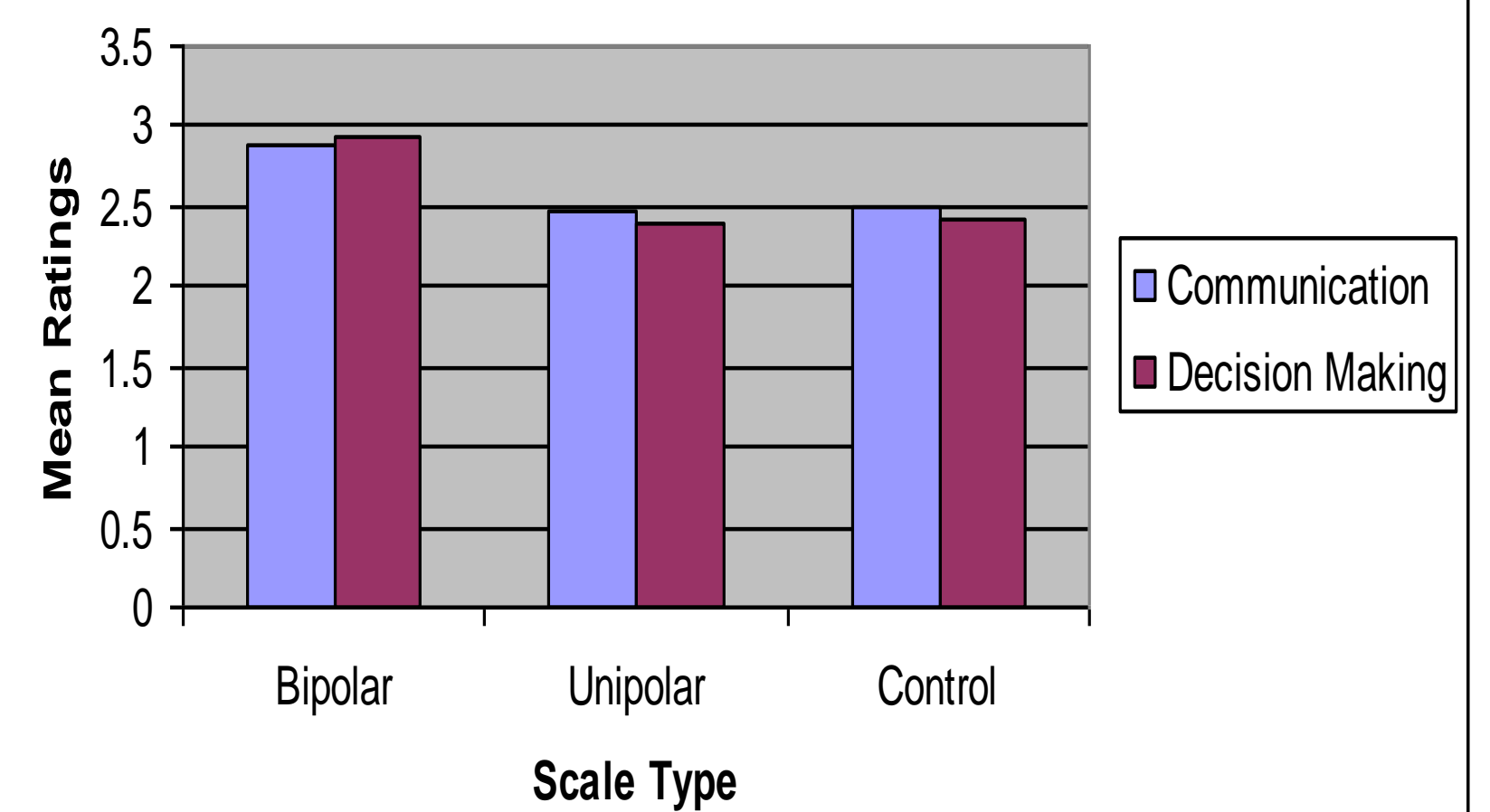
Discussion

• The current results suggest that participants responded more leniently when using a bipolar rating scale than when using either a unipolar rating scale or a control scale without anchor labels. It is clear from these findings that scale polarity and associated anchor label wording may affect the accuracy of performance ratings.

• Organizations are using employee performance data in myriad ways, including decisions for workforce planning and staffing, compensation, and training and development. Sound organizational decisions regarding workforce talent management require the highest possible level of accuracy and objectivity in employee performance measurement and interpretation to achieve business objectives.

References

• Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey international consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement, 60*, 361–370.
• Horan, P. M., DiStefano, C., & Motl, R. W. (2003). Wording effects in self-esteem scales: Methodological artifact or response style? *Structural Equation Modeling, 10*, 435–455.
• Mazaheri, M., & Theuns, P. (2009). Effects of varying response formats on self-ratings of life-satisfaction. *Social Indicators Research.* 90, 381-395.
• Tourangeau, R., Couper, M.P., Conrad, F. (2007). Color, labels, and interpretive heuristics for response scales. *Public Opinion Quarterly. 71(1),* 91-112.

Table 1: Ratee 1 Communication Performance Dimension

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4.848	2	2.424	3.077	.050
Within Groups	98.466	125	.788		
Total	103.314	127			

Table 2: Ratee 1 Decision Making Performance Dimension

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7.707	2	3.853	4.542	.012
Within Groups	106.048	125	.848		
Total	113.755	127			