



# The Effect of Rater Consensus on Performance Rating Accuracy



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

## Abstract

This study examined the accuracy of performance ratings provided by participant raters with and without a consensus requirement. Participants in three conditions, discussion with consensus, discussion without consensus, and a no discussion control condition, evaluated the performance of three ratees working on a problem solving exercise. It was hypothesized that ratings provided by participants in the consensus condition would yield greater accuracy than participants' ratings in the other conditions. Findings in support of this hypothesis offer justification for use of multiple raters reaching consensus in organizational performance appraisal situations.

## Introduction

- A performance management strategy is comprised of assessment, feedback, and reactions, with each area consisting of specific strategies. A well-designed and implemented performance management system offers value in terms of a high-performing workforce and achievement of organizational objectives (London, Mone, & Scott, 2004).
- Researchers interested in the feedback aspect of performance management have generally focused on two constructs:
  - Behavioral accuracy: correct identification of whether a behavior occurred
  - Rating accuracy: appropriate rating of a behavior and extent to which it matches a standardized rating score
- Cronbach (1955) developed four measures to examine rating accuracy. Each measure requires rating scores provided by designated raters (i.e., observed scores) and standardized rating scores provided by one or more trained expert raters (i.e., true scores).
  - Elevation (E): the accuracy of the average rating across all ratees and dimensions provided by a rater
  - Differential elevation (DE): the accuracy of the average rating given to each ratee across job dimensions
  - Stereotype accuracy (SA): the accuracy of the average rating given to each job dimension across ratees
  - Differential accuracy (DA): the accuracy on a specific ratee and a specific performance dimension

- Performance research has examined the extent to which different types and number of ratees, performance criteria, and contextual factors impact rating accuracy (Tetlock, 1985; Salvemini, Reilly, & Smither, 1993). However, few research studies have examined the extent to which rating consensus through the use of multiple raters in collaboration may increase rating accuracy. Roch (2006) examined the extent to which group discussion and consensus affect rating accuracy. Findings showed significant improvements in rating accuracy after reaching consensus.
- The focus of this study is on the implications of multi-rater discussion and consensus on rating accuracy. It is hypothesized that participants in the consensus condition will have greater rating accuracy than participants in the other conditions.

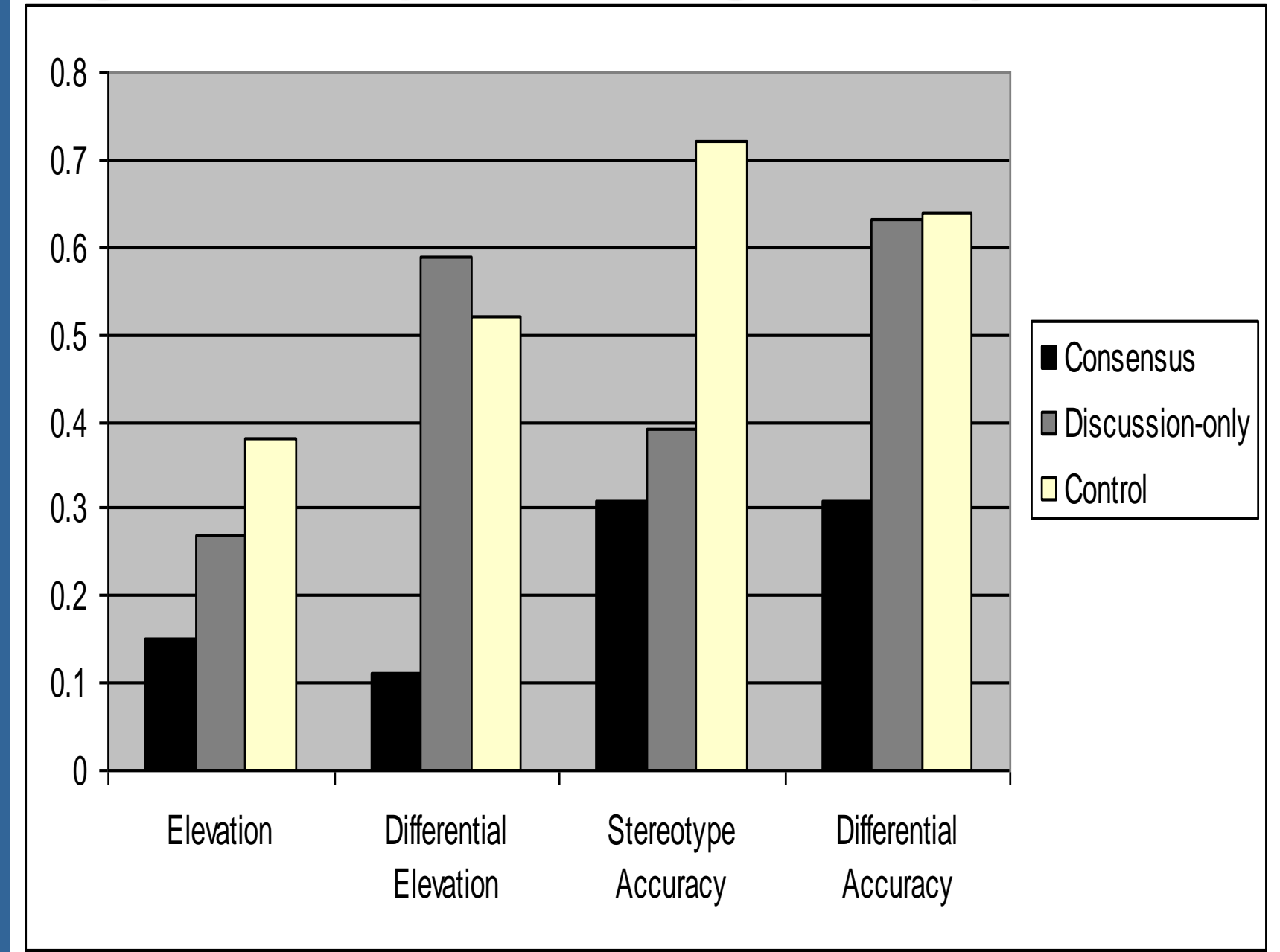
## Method

- Participants (n=75) were randomly assigned in groups of three to one of the three conditions. Groups viewed a video depicting three ratees working on a problem solving exercise. Instructions for use of a 7-point rating scale and definitions of the three performance dimensions (verbal communication, collaboration, and problem solving) were provided to all groups.
- The participants evaluated each ratee on demonstrated behaviors within the performance dimensions. Following rating completion, participants were debriefed and adjourned.
- Cronbach's (1955) indexes provided four measures for determining extent of variance between participant rating scores (i.e., observed scores) and previously established expert rating scores (i.e., true scores).

## References

- Borman, W.C. (1977). Consistency of rating accuracy and rating errors in the judgment of human performance. *Organizational Behavior and Human Performance*, 20, 238-252.
- Cronbach, L.J. (1955). Processes affecting scores on understanding of others and assumed "similarity." *Psychological Bulletin*, 52, 177-193.
- London, M., Mone, E.M., & Scott, J.C. (Winter 2004). Performance management and assessment: Methods for improved rater accuracy and employee goal setting. *Human Resource Management*, 43(4), 319-336.
- Roch, S.G. (2006). Discussion and consensus in rater groups: Implications for behavioral and rating accuracy. *Human Performance*, 19(2), 91-115.
- Salvemini, N.J., Reilly, R.R., & Smither, J.W. (1993). The influence of rater motivation on assimilation effects and accuracy in performance ratings. *Organizational Behavior and Human Decision Processes*, 55, 41-60.
- Tetlock, P.E. (1985). Accountability: The neglected social context of judgment and choice. In B.M. Staw & L. Cummings (Eds.), *Research in Organizational Behavior* (Vol. 7, 297-332). Greenwich, CT: JAI Press.

Figure 1: Group Level Rating Accuracy Means



## Results

- A series of one-way ANCOVA calculations yielded significant differences between scores for each of the four Cronbach (1955) accuracy indexes. The consensus condition demonstrated significantly higher rating accuracy for each measure of accuracy than the other two conditions overall.
- Findings provided support for the hypothesis (p<.05) that participants in the consensus condition would demonstrate a greater degree of rating accuracy than participants in the discussion without consensus and control conditions.

### Elevation

Results indicated a significant difference between scores across all conditions,  $F(2, 74) = 3.491, p < .05$ .

### Differential Elevation

Results indicated a significant difference between scores across all conditions,  $F(2, 76) = 2.812, p < .05$ .

### Stereotype Accuracy

Results indicated a significant difference between scores across all conditions,  $F(2, 74) = 2.377, p < .05$ .

### Differential Accuracy

Results indicated a significant difference between scores across all conditions,  $F(2, 75) = 4.801, p < .05$ .

## Discussion

- The performance evaluation process must become more adaptable in response to increasingly complex jobs, greater incumbent interaction and collaboration, and shifting organizational objectives.
- Though the use of multi-rater feedback in performance management is increasing among organizations, critical issues exist regarding proper implementation and use. A consensus requirement for a multiple rater scenario, along with appropriate rating accuracy measures, may offer a feasible method for improvement and enhanced value.
- The present research study attempts to provide insight into the extent to which a consensus-driven performance rating model may improve the accuracy of performance ratings in the context of two relevant organizational factors, multiple ratees and multiple job-specific performance dimensions.