

# Developments In Business Simulation & Experiential Exercises, Volume 23, 1996

## PILOT ANALYSES OF SELF-PEER EVALUATIONS IN AN EXPERIENTIAL-EXERCISE HUMAN RESOURCES MANAGEMENT COURSE

John D. Overby, The University of Tennessee at Martin  
Festus Olorunniwo, The University of Tennessee at Martin

### ABSTRACT

The reported study is a pilot analyses consisting of four performance concepts and four Human Resource Management theoretical concepts. Sixteen measures were analyzed to determine if there was a significant difference between Self and Peer performance evaluations. These concepts and evaluations were based on the performance of fifteen experiential exercises related to fifteen theoretical concepts and Self-Peer evaluated on eleven performance concepts. Results indicate that Peer and Self-ratings may be significantly different across the myriad of 165 interaction measures. The theoretical concept of Job Analysis was found to be significantly different across all four-performance measures. The remaining three theoretical concepts were found to contain only one or two significant differences.

### INTRODUCTION

Use of peer and self-peer evaluations have been a subject of interest, not only in the academic settings, but also in the industrial settings. In the classroom, however, it is not a common practice to use self or peer evaluations. In the context of the study reported here, the first author conjectured that there is a likelihood of a tendency for students to overrate themselves to increase their own scores.

Compared with self-evaluations, peer evaluations could be less or greater than self-evaluation depending on a range of factors. Such factors may include the reluctance to underrate a peer, because of the fear that a "peer" may do the same to you. Other factors may depend on the personality traits of the peer evaluator (low/high self-esteem; extraversion/introversion; agreeableness/disagreeableness, etc.) While the first author had a definite conjecture regarding self-evaluation, the existence of a myriad of factors contributing to peer-evaluations relative to self necessitates that the issue be studied in the classroom setting.

Use of peer and self-peer evaluations has been studied in a multitude of academic and industrial settings as a means to determine the performance appraisal of work assignments. The industrial settings include research by Harris and Schaubroeck (1988) and Cooke, et al (1987). Harris and

Schaubroeck (1988) found in a meta-analysis of the literature a relatively high correlation between peer and supervisor ratings, but only a moderate correlation between self-supervisor and self-peer rating, with the type of job seeming to be a moderator of the self-peer and self-supervisor ratings. Cooke, et al, (1987) found through factor analysis that coworkers generally agreed with each other's ratings, but correlations between self and peer rating were lower.

Within the academic settings, peer and self-peer ratings have been researched in a wide variety of disciplines. Alagna and Reddy (1985) studied the use of self and peer evaluations in anatomy dissection. Their analyses found significant differences between mixed-sex and male groups peer-evaluations in that male groups tended to rate higher than mixed-sex groups. Marcoulides and Simkin (1991) researched within business administration education and their findings indicate that students can be consistent and fair in their assessments. McKendy (1990) found in a composition course that students are consistent in their repeated evaluations of an essay.

Without specification of the academic discipline, much additional research has been conducted. Fry (1990) found students believed their work had been graded fairly by peers and that the grades should be included in their final course grade. John and Robins (1993) found agreement in self-peer evaluations was highest for traits related to Extraversion and lowest for traits related to Agreeableness. Their research also showed that on average, self-peer agreement was lower than peer-peer agreement. Watson and Clark (1991) found that self-peer agreement increased with the addition of more peer raters and with greater peer acquaintance. O'Connor and Day (1989) found that there was more variance in peer ratings than in self-ratings. Saaverdra and Kwun (1993) found that outstanding contributors were the most discriminating evaluators and that self-evaluations were higher than peer evaluations. Keller and Bishop (1985) found that self-esteem and affinity had a significant effect on peer rating and suggested that ratings by peers high in self-esteem may have been more bias-free than raters with low self-esteem. Raters with low self-esteem tended to impose excessively rigorous standards in their evaluations.

# Developments In Business Simulation & Experiential Exercises, Volume 23, 1996

## THE STUDY

Though the research of self-peer evaluations is rich in findings relative to perceptions and personal-attributes, little research has been conducted to determine the impact of specific performance concepts and the variance associated with specific theoretical concepts of an academic discipline. The present study sought to analyze the significance of specific performance concepts and specific theoretical concepts set within the discipline of human resource management and the performance of experiential exercises. The study was conducted using ten specific performance concepts plus an overall assessment, and fifteen theoretical concepts within the discipline across a fifteen-week semester. The ten specific performance concepts that the students used as a basis for exercise performance evaluation were specific to the text used for the course and are listed in Table 1. Students were also provided a definition in their text of the ten specific performance concepts.

Each performance is evaluated on a seven point scale ranging from: 1 )poor, 2)well below satisfactory, 3)below satisfactory, 4)satisfactory, 5)above average, 6)very good, 7)outstanding, plus the opinion of not observed.

TABLE 1

TEN SPECIFIC PERFORMANCE  
CONCEPTS STUDIED

Analytic Thinking
Behavioral Flexibility
Decision Making
Leadership
Oral Communication and Presentation
Perception of Threshold Clues
Personal Impact
Planning and Organizing
Self Objectivity
Written Communication
Overall Assessment

The 15 human resource management (HRM) theoretical concepts are the 15 chapters covered during the semester and are listed in Table 2.

TABLE 2

FIFTEEN THEORETICAL CONCEPTS STUDIED  
HRM CONCEPT                      EXPERIENTIAL  
EXERCISE TOPIC

HRM in a Changing Environment	Assessment of Customer Satisfaction and Relationship to HRM
Competitive Advantage through HRM	HRM Issues affecting Competitiveness
Equal Employment Opportunity	Selection Procedures and Race Discrimination
Job analysis	Critical Incident Technique & University Professors
HR Planning and Recruitment	Turnover as a Problem in HRP
Personnel Selection	Use of the Wonderlic Test and Adverse Impact
Employment Interviews	An Evaluation of Interview Questions
Organizational Training	Development and Evaluation of training for Graduate Instructors
Career Development	Career Development Self-Assessment Exercise
Performance Appraisal	Development of Performance Appraisal System for Instructors
Direct and Indirect Compensation	Factors Related to Pay Determination
Pay for Performance	Design of a Pay Determination
Improving Quality, Productivity, QWL	Design of a Pay for Performance System
Labor Relations	Attitudes toward Unions
Employee Health and Safety	Development of a Company Smoking Policy

# Developments In Business Simulation & Experiential Exercises, Volume 23, 1996

## PROCEDURES

One class meeting per week was devoted to the performance of an experiential exercise. For each exercise the students were required to do some “home-work prior to arriving at class. The students were then randomly assigned to one of four to five groups consisting of five to six students, dependent on the size of the class and the attendance the day of the exercise. The group members then arranged into a circular form and a group spokesperson was chosen or volunteers to present the group’s views in the closure discussion at the end of the class period. The group members then read each peers “homework” before the group’s began a discussion. The discussion amongst each group’s members continued until a consensus was derived. Once the groups derived a consensus, the instructor surveys the groups’ consensus and conducts a closure discussion. Once the closure discussion concludes, the students were instructed to evaluate Self and the Peer seated to their left or right and to turn in their “homework” and the two performance evaluations to the group spokesperson who in turn transfers the materials to the instructor.

Data for the study was collected from four class sections over four 15-week semesters. A total of 108 students’ performance evaluation scores are included in the study’s data set. For each of the 11 performance concepts and each experiential exercise, each student contributes a self and a peer evaluation score. Thus, for each of the 108 students, providing two evaluations each for everyone of the 15 exercises over all the 11 performance concepts, there are 330 cells of data. Thus, the total data set is composed of 35,640 observations. It should be noted that the students may choose to not participate in two of the experiential exercises if s/he so chooses and thus some of the cells contain missing scores.

## TESTS OF THE HYPOTHESES

In order to reduce the data matrix to a manageable size, the difference:  $d = \text{peer score} - \text{self-score}$ , was calculated for every observation. Using this difference, the following hypotheses were tested:

$$H_0: d_{ij} = 0, i = 1,2,3,\dots,15; j = 1,2,3,\dots,11$$

(peer minus self evaluations score = 0)

$$H_1: d_{ij} > 0, (\text{peer minus self score} > 0)$$

The paired difference  $d_{ij} = p_{ij} - s_{ij}$   
where

$p_{ij}$  = Peer evaluation for exercise  $i$  ( $i = 1,2,3,\dots,11$ ) tested against performance concept  $j$  ( $j=1,2,3,\dots,15$ )

$s_{ij}$  = self evaluation for exercise  $i$  ( $i = 1,2,3,\dots,11$ ) tested against performance concept  $j$  ( $j = 1,2,3,\dots,11$ )

Note that the alternative hypotheses  $H_0$  and  $H_1$  are equivalent to the hypotheses:

$$H_0': p_{ij} = s_{ij} \text{ (peer = self evaluation score)}$$

$$H_1': p_0 > s_{ij} \text{ (peer > self evaluation score)}$$

For example, by fixing  $i$  at  $i=1$ (HRM in Changing environment), we tested if there is a significant difference in peer minus self-evaluation for each concept. Thus by repeating the same procedure for all the 15 exercises, it is possible to generate a 15x11 matrix of p-values and  $n$  (sample size) values.

This study presents a pilot study of a limited subset of the full matrix. Specifically, we chose the subsets of performance concepts and theoretical concepts shown in Table 3. These eight concepts were chosen because the authors felt that they were the most relevant to the subjects of analysis and evaluation, in general, or were a topic of interest in current management literature.

TABLE 3

PERFORMANCE AND THEORETICAL CONCEPTS OF PILOT STUDY

Performance Concepts	Theoretical Concepts
Analytical Thinking #1	Equal Employment Opportunity #3
Leadership #4	Job Analysis #4
Self Objectivity #9	Performance Appraisal #9
Overall Assessment #11	Improving Quality #13

## SUMMARY OF THE FINDINGS

Analyses of the 16-cell matrix, shown in Table 4, indicate that across the four performance and the four theoretical concepts, nine significant differences

## Developments In Business Simulation & Experiential Exercises, Volume 23, 1996

were present between the Peer and Self evaluations with  $p < .10$ . These significant differences indicate that Self tends to rate their performance lower than Peer ratings. The theoretical concept of Equal Employment Opportunity had one significant difference within the performance concept of Overall Assessment. Across the theoretical concept Job Analysis significant differences were found in all four of the performance concepts. The Performance Appraisal theoretical concept possessed two significant differences within the performance concepts Leadership and Overall Assessment. Significant differences across the theoretical concept, Quality Improvement, were found in the two performance concepts of Analytic Thinking and Self-Objectivity.

Syntheses of the findings indicate that Peer and Self as evaluators are discriminating in their deliberations of performance by themselves and peers. The synthesis also indicate that the significant differences occur in a varying fashion across the four performance concepts. In only one case, Job Analysis, were all four-performance concepts significantly different. Our final finding is that Self tends to evaluate their performance equal to or lower than Peer evaluates Self's performance.

### CONCLUSIONS AND RECOMMENDATIONS

Results of the pilot study indicate that Self and Peer can be significantly different in the performance discriminations of experiential exercises. These findings indicate that Self tends to rate their own performance below the ratings of Peer.

Findings also indicate that certain theoretical concepts possess more significant differences than other theoretical concepts across the performance concepts studied. Thus it appears that there is adequate reason to expand the study to a complete analyses of the data to determine the fullest extent of the significant difference in performance ratings across the fifteen theoretical concepts and eleven performance concepts.

The study should also be expanded to consider other intervening variables such as demographic variables. Demographic variables that should be considered include, but should not be limited to, gender, major, grade point average, race, seasonality (semester).

### REFERENCES

- Alagna, Sheryle W. and Diane M. Reddy (1985) Self and Peer Ratings and Evaluations of Group Process in Mixed-Sex and Male Medical Training Groups. Journal of Applied Social Psychology, 15, 31-45
- Cooke, Robert A. et al. (1987) Thinking and Behavioral Styles: Consistency Between Self-Descriptions and Descriptions by Others Educational and Psychological Measurement, 47, 815-823.
- Fry, Stuart A. (1990). Implementation and Evaluation of Peer Marking in Higher Education. Assessment and Evaluation in Higher Education, 1 5, 177-189.
- Harris, Michael M. and John Schaubroeck (1987) A Meta-Analysis of Self-Supervisor, Self-Peer, and Peer-Supervisor Ratings. Personnel Psychology, 41, 43-62.
- John, Oliver and Richard Robins (1993) Determinants of Interjudge Agreement on Personality Traits: The Big Five Domains, Observability, Evaluativeness, and the Unique Perspective of the Self. Journal of Personality, 61, 521-551.
- Keller, Betty G. and Ronald C. Bishop (1985) Self-Esteem as a Source of Raters' Bias in Peer Evaluation Psychological Reports, 56, 995-1000.
- Marcoulides, George and Mark G. Simkin (1991) Evaluating Student Papers: The Case for Peer Review. Journal of Education for Business, 67, 80-83.
- McKendy, Thomas F. (1990) Legitimizing Peer Response: A Recycling Project for Placement Essays. College Composition and Communication, 41, 89-91
- O'Connor, Brian P. and Russell Day (1989) Could Differences in the Interpretation of Behavior be a Reason for Low Rating-Behavior, Self-Peer, and Peer-Peer Personality Correlations? Social Behavior and Personality, 17, 17-28.
- Saaverdra, Richard and Seog K. Kwun (1991) Self-versus Peer Ratings of Specific Emotional Traits: Evidence of Convergent and Discriminant Validity. Journal of Personality and Social Psychology, 60, 927-940.

**Developments In Business Simulation & Experiential Exercises, Volume 23, 1996**

**TABLE 4  
P-VALUES AND POPULATION SIZE  
OF PILOT STUDY PERFORMANCE AND THEORETICAL CONCEPTS**

**Performance Concepts**

Theoretical Concepts	Analytical Thinking	Leadership	Self-Objectivity	Overall Assessment
EEO	p = .8564 n = 86	p = .6959 n = 87	p = .4206 n = 83	p = .0384* n = 71
Job Analysis	p = .0129* n = 75	p = .0243* n = 75	p = .0572* n = 74	p = .0101* n = 71
Performance Appraisal	p = .2183 n = 87	p = .0422* n = 89	p = .2177 n = 89	p = .0102* n = 80
Quality Improvement	p = .0584* n = 90	p = .2611 n = 90	p = .0617* n = 90	p = .2884 n = 84

\* Significance if p < .10