A COMPARISON OF SELF VERSUS PEER EVALUATIONS IN TEAM PROJECTS USING ALTERNATIVE SCALING FORMATS: IMPLICATIONS AND ISSUES

Stuart Van Auken
California State University, Chico
Department of Finance & Marketing
Chico, CA 95929-0051
(530) 898-5666

ABSTRACT

Self evaluations as well as peer evaluations are typically used to evaluate team-member performance in group projects. This study is concerned with the extent of inflation in self responses relative to peer evaluations on three traits over three scaling formats. The results suggested that the ratio scale produced the greatest inflation followed by the semantic differential scale. Further, a scale anchored to expectations and which was designed to reveal the social loafer and the supra normal performer had the least inflation, as the magnitude of the scale intervals was so large that inflated responses were inhibited.

INTRODUCTION

Team projects in marketing courses continue to receive scrutiny (Beatty, Haas, and Sciglimpaglia 1996; Dommeyer 1966; Williams, Beard, and Rymer 1991) and this attention is necessary given the need for the development of team-building skills among business students (Wright, Bilner, and Zeithaml 1994). Some of this scrutiny has involved the use of peer evaluations with a focus on the assignment of grades in an equitable manner (Beatty, Haas, and Sciglimpaglia 1996). Additional research has also involved a comparison of alternative scaling formats in the assessment of team-member performance (Van Auken 1994 and 1995). In this regard, how do alternative scales perform in the revelation of the social loafer and supra normal performer? Further, a plethora of controls and procedures has been developed to evaluate individual team-member performance (Williams, Beard, and Rymer 1991). Most notably, the use of triangulation.

Despite these enhancements in evaluations, one area that has not been scrutinized is the self evaluation of a team-member’s performance relative to the evaluation of other team members. In turn, a number of questions are raised. To illustrate, are self evaluations inflated? Are they honest? Should they be discarded? Do they have a meaningful role to play in team-building projects?

On the surface, self evaluations help to combat anomie by giving students a sense of empowerment. In essence, a student may perceive a self evaluation as a means of defending oneself, especially if personality conflicts exist among team members. Self evaluations may also allow one to display a situational self image (Schenk and Holman 1980). That is, a perception of one’s performance image relative to the perception of others. However, for some, a self evaluation may be defined as a “looking glass self” (Sirgy 1982), or the performance image that one believes that others hold. By assessing the relationship between self evaluations and fellow team-member evaluations in team projects, suggestive and cursory insights can be achieved as to which self is being activated.

Questions of honesty also abound with respect to self evaluations. For many students, the pressure to receive good grades results in the need to take advantage of any opportunity that is presented to them and this could entail inflated self evaluations (Roberts and Rabinowitz 1992). Basically, academically dishonest behavior is a way of dealing with academic pressures (Todd-Mancillas and Sisson 1987). Further, recent studies suggest that the propensity and magnitude of cheating by business majors is both widespread and increasing (Meade 1992).

To achieve insights into these issues, one must determine if there are meaningful discrepancies between individual self perceptions and the perceptions of others. If so, a number of hypotheses may be developed. This study will therefore operate in an exploratory sense to determine the relationship between average self-
evaluation scores on each of three traits relative to average peer evaluation scores on the same three traits across three alternative scaling formats. The latter, of course, is designed to reveal scale sensitivity, as the nature of a scale could impact self evaluations based on its anchoring.

**THE STUDY**

In an effort to generate the data for the study, ten student teams comprised of five students each who had completed a marketing research project provided individual self evaluations and peer evaluations on the following three traits: one’s extent, quality and overall contribution to the assigned project. These traits were measured on each of three measurement scales: ratio, semantic differential and a newer scale entitled “meets expectations.” These scales were systematically rotated among individual team members for each student group to help control for order bias.

The ratio scale appears in Figure 1 and its purpose is to achieve relative insights into individual team member performance on each of the three criteria. The semantic differential scale appears in Figure 2 and it replicates the traits utilized by Haas and Sciglimpaglia (1994). Among these traits are variables that encompass the extent, quality, and overall nature of one’s contribution. These traits were selected for analysis because they were common to both the ratio and the to be discussed “meets expectations” scale. The semantic differential scale is unique in that each group member is evaluated in an absolute way and not in a relative sense on each of the indicated traits. In other words, each team member is not directly compared against other team members as is the situation for the ratio scale. Finally, the “meets expectations” scale appears in Figure 3. In this scaling format, individual team members are evaluated on each of the three indicated traits, yet these traits are anchored to group expectations which are formed at the initiation of a project (Van Auken 1996). This scale is unique in that it forces evaluators to relate each peer and themselves to a group performance norm. Although team members are being evaluated independently, like in the semantic differential scale, the forced comparison with the central scale position results in a judgment as to who meets and does not meet group expectations.

As can be observed the scales are very different. The ratio scales provides relative judgments, while the semantic differential scale produces absolute judgments. The “meets expectations” scale produces absolute judgments, yet the anchoring approach is similar to the conceptual effect of the constant sum scale. It is these scale differences that are of interest in assessments of self versus peer evaluations. Basically, the “meets expectations” scale may not encourage inflated responses to self evaluations as all respondents are related to a group performance norm. Deviations from the norm are thus quite conspicuous.

Given a total of fifty students, 50 self evaluations were obtained as well as 200 peer member evaluations (i.e., 20 peer evaluations per student team). In turn, two groups were created: a self evaluation group and a peer evaluation group. An assessment of the differences between these groups on the three traits across the three scaling formats thus represents the heart of this study.

**THE RESULTS**

Table 1 portrays the mean response for each trait and the accompanying Z score for each of the alternative scaling approaches. As can be determined from the results of one-way ANOVA tests, two of the three scaling formats evidence statistically significant differences between self and peer evaluations. In this regard, the semantic differential scale evidenced a significant difference on all three traits, while the ratio scale revealed a difference on two traits with a near miss on the quality trait (p < .06). Noteworthy is the lack of a statistically significant difference for each of the three traits measured by the “meets expectations” scale.

As can be further observed from Table 1, the mean scores on every trait across the three scaling formats were higher for self evaluations when contrasted to peer evaluations. To help dramatize this, the data were standardized to a mean of zero and unit variance. The resulting Z scores thus show the extent that a given trait mean is over or below the sample average. In this regard, positive Z scores denote above average responses, while negative Z scores evidence below average responses. The higher the absolute magnitude of a Z score, the greater its distance from the overall sample mean. As can be noted in Table 1, the
overall contribution trait using the ratio scale produced a Z value of .402 for the self evaluation and a value of -.100 for the peer evaluation. This was the highest mean separation and the lowest probability of committing an alpha error (P < .00).

A Multiple Discriminant Analysis was also run on each of the scaling formats and the percentage of correct classifications for each of the scaling formats are as follows: ratio = 70.4%, semantic differential = 50.8%, "meets expectations" = 63.6%. Given the lack of a statistically significant difference on any trait in the latter scaling format, the discriminant function is viewed as spurious and the resulting classification is due to chance. Overall, the ratio scale's discriminant function produced a classification percentage which exceeded the proportional chance criterion of 68.0%, while the semantic differential scale did not. Thus, the ratio scale invites the greatest inflation in self evaluation, followed by the semantic differential scale, while the "meets expectations" scale is relatively clean.

Apparently, judgments that force relative evaluations result in greater inflation in self evaluations and this may be a caveat associated with the ratio scale. The semantic differential scale also invites inflation in self evaluations; however, it does not force relative judgments. Finally, the "meets expectations" scale does not invite inflation on the magnitude of the other scales as it is anchored to a meeting of expectations and departures from this norm are far more conspicuous.

IMPLICATIONS

The results suggest that inflated evaluations of the self in team projects do exist and that a situational self image involving one's performance image relative to others does exist. It is significant to note that this image is apparently activated when relative judgments are made or when there is an expectation of an instructor comparison among peers. The results also suggest that the revealed inflation may be somewhat academically dishonest. Further, one can conjecture that the western value of individualism and it tenets of success and winning may be a factor in this dishonesty (Triandis 1985). Basically, the instructor may invite dishonesty depending upon the selection of a scaling format for team member evaluations. Overall the cleanest scale is the "meets expectations" scale. It can reveal the social loafer and the supra normal performer and has the added benefit of forcing a discussion of expectations among all team members prior to the initiation of a team project.

ISSUES

The actual project that students worked on and supplied the indicated team member evaluations was a difficult one and there was variation among and within teams as to performance. It is therefore possible that simpler projects will not result in inflated self perceptions as to the extent found here. This project also did not involve the use of triangulation as a means of evaluation. In essence, the use of multiple evaluation strategies may inhibit inflated self evaluations. Some of these multiple methods encompass confidential memos, instructor observation, and interaction logs (Williams et al. 1991).

This study was also exploratory, yet it does permit the creation of a number of hypotheses as differences were found between self and peer evaluations. One hypothesis is that situational image is being measured relative to "looking glass" image as the latter would not invite discrepancies between self and peer evaluations.

Another hypothesis is that the nature of the measurement scale invites inflated self evaluations and this may be particularly true when relative evaluations are being sought. However, one may also hypothesize that absolute judgments, as found in the semantic differential scale, will generate inflated self evaluations; especially, if students perceive that the instructor will make relative comparisons among team members.

A final hypothesis would be that a "meets expectations" scale would lessen inflated self evaluations relative to one's peers as a departure from a meeting of expectations would produce a more conspicuous result. Finally, a replication of this study under conditions of triangulation versus nontriangulation may further develop a theory for peer and self evaluation measurement.

CONCLUSIONS

This exploratory study suggests that inflated self evaluations in the assessment of team-member performance may be a function of the measurement
scale and that such evaluations may be influenced by a lack of triangular controls in team evaluations. Although these inflations may be dishonest, they would appear to occur when students are given scales that require relative judgments, or when there is a perception among students that the results will be used in relative student evaluations. Uniquely, a scale anchored to expectations lessened inflated self evaluations as team-member performance was related to group norms for the extent, quality and overall nature of one's contribution. Deviations from these norms are possibly of a higher psychological magnitude than the inflated self perceptions of ratio and semantic differential scales, and may call attention to oneself in a way that is unlike the other scaling approaches. Thus, the "meets expectations" scale may inhibit an inflated response. All in all, self evaluations combat a sense of powerlessness among team members and would appear to function best in an environment characterized by multiple controls. Finally, the "meets expectations" scale tends to discourage a dishonest response.

REFERENCES


Figures 1-3 and Table 1 are available from the author by request.