

WHEN MORE IS MORE: CAN THE BALANCED SCORECARD MITIGATE THE IMPACT OF STEREOTYPING ON PERFORMANCE APPRAISALS?

Branson, Leonard

University of Illinois Springfield

Steele, Nathan

University of Illinois Springfield

Sung, Chung-Hsien

University of Illinois Springfield

ABSTRACT

An organization's ability to accurately identify and reward managers moving it toward its strategic objectives is crucial. The accuracy of employee performance evaluation systems continues to be problematic. Stereotyping has proven to be a very robust negative phenomenon that both researchers and practitioners strive to control. Financial performance has been the gold standard in determining if organizations and their managers were performing well. The introduction of the Balanced Scorecard changed the framework from looking primarily at financial performance, to looking at a balanced set of financial and non-financial performance indicators. This paper reports an empirical study examining the impact of stereotyping on the performance ratings of managers in either the context of only financial performance measures, or the context of using BSC performance measures. Our results indicate that stereotyping has an impact on evaluations in both contexts, but the use of the BSC significantly reduces the impact of stereotype information on managerial performance evaluations below levels observed when solely utilizing financial performance data.

INTRODUCTION

Both researchers and practitioners have long struggled to try to understand and control the unwelcomed influence of non-performance factors on performance evaluations (Spence & Keeping, 2010). Heneman (1986) conducted a metaanalysis of studies which reported on the relationship between ratings and results. He concluded that the relationship between results and ratings is relatively weak, and that in some situations there is virtually no relationship, or even a negative relationship between results, behaviors, and ratings.

Although the argument of whether it is better to evaluate and reward based on effort or results has not been settled, most agree that for mid- and upper-management, results are very important. Markets hold companies and their leaders responsible for results, and market influences cascade downward in a company putting pressure on management at all levels to meet market expectations in results. However, previous research is unambiguous in its

findings that performance ratings are heavily influenced by non-performance factors (Heneman, 1986, Branson, Moe, & Sung, 2005, Branson, Steele, & Sung, 2010a, Steele, Branson, & Sung, 2011, Bol & Smith, 2011) and not just by results.

A fundamental problem with performance ratings is the limited information processing capacity of human decision makers. Humans can process very little information at a time, and we make extensive use of cognitive categories (schema) and heuristics in making decisions (Taylor & Crocker, 1981). A schema is a generic term for a cognitive category, or knowledge structure, which guides information processing (Taylor & Crocker, 1981). Stereotypes and implicit personality theories are specific examples of schemata. Whatever their nature or type, schemata are seen as controlling the selectivity that occurs in each of the information processing stages of perception, encoding, storage, retrieval, processing and evaluation. Unfortunately, although schematic processing is an efficient mechanism, it can lead to systematic biases in perception, recall, processing and decision making. Nonrepresentative information or judgment-irrelevant information often influences schematic processing and produces biased conclusions, even though schemata are inherent in human cognition (Hastie, 1981; Taylor & Crocker, 1981; Steele, Branson, & Sung, 2011).

Since Landy and Farr (1980; 1983) there has been support for the need to incorporate rater cognitive complexity into our understanding of the appraisal process. Feldman (1981) and Ilgen and Feldman (1983) developed a model which recognizes that most appraisers are faced simultaneously with multiple cognitive tasks. High cognitive load and stress forces suboptimal evaluations. There is substantial evidence that appraisals are biased by the categorization process. Feldman (1981) maintains that the prototype or schema to which a person is assigned will direct future information processing about that person. The characteristics of the prototype are recalled, and these characteristics are attributed to the evaluatee, whether or not they are real.

While there are numerous streams of research in performance management, the most widely applied method of using a multiple source, multiple measure approach to performance management was introduced by Kaplan and Norton (1992): the Balanced Scorecard (BSC). The BSC proposes the use of a matrix of objective outcome measures as the source of information for performance evaluation decisions. Nevertheless, the use of subjective measures and a variety of other problems continues to make its influence felt even in the BSC context (Ittner, Larcker, & Meyer, 2003; Cardinaels, & van Veen-Dirks, 2010; Roberts, Albright, & Hibbets, 2004; Wong-On-Wing, Gue, Wei, & Yang, 2007). This paper examines the efficacy of the BSC in reducing the influence of subjective, performance irrelevant information on performance ratings.

LITERATURE REVIEW

There are many streams of research in performance appraisal in psychology, social psychology, industrial psychology, and in the more applied areas of business such as management and accounting. Previous research has investigated individual differences within and between raters and ratees including demographic differences (Wendelken & Inn, 1981), cognitive differences (Feldman, 1981), and social intelligence differences (Branson, Steele, & Sung 2011). Demographic differences that have been found to influence performance ratings include age, race, and gender differences between ratees (Wendelken & Inn, 1981) as well as the raters (Wendelken & Inn, 1981), and the interaction between age, race and gender of ratee and rater (Wendelken & Inn, 1981). Wang, Wong and Kwong (2010) found that raters with harmony, fairness and motivating goals inflated their ratings for low performers more than for moderate or high performers, and they deflated the ratings of high performers if their goal was fairness. Another important individual difference is the emotional competency and social intelligence of the rater (Branson, Steele, & Sung, 2011) and of the ratee (Kim, Cable, Kim & Wang, 2009).

Organizational factors that have been investigated include the purpose of the rating (raise, promotion or developmental review; van Veen-Dirks, 2010), if the process is conducted by individuals or groups (Branson, Steele, & Sung, 2010a), if the process is conducted by groups with different characteristics (Branson, Steele, & Sung, 2010b), if those groups were virtual or collocated (Branson, Moe, & Sung, 2005; Steele, 2010). Other factors include the number, types, diversity, and source of performance measures used (Moers, 2005; Branson, Steele, & Sung, 2010a), and whether the appraisal process is seen as a motivational, communication or socialization process (Findley, Giles, & Mossholder, 2000).

The purpose (developmental versus administrative) of the rating and the rating sources introduce systematic variances in the rating values (Greguras & Robie & Schleicher & Goff, 2003; van Veen-Dirks, 2010). While companies have long used multiple measurements for employee development (identifying strengths and weaknesses, training), the use of multiple source measurements for administrative (promotion, raises, etc) purposes gained wide spread use with the implementation of the Balanced Scorecard that was introduced by Kaplan and Norton (1992). Subordinate ratings based on multiple source measurements are significantly better for developmental than administrative purposes (Greguras & Robie & Schleicher & Goff, 2003). Ittner, Larker and Randall (2003) found that greater measurement diversity and better alignment of strategy with firm value drivers resulted in higher measurement system satisfaction and better stock market performance.

Krzystofiak et al. (1988) found that since appraisal has a person focus, raters tend to use person schema for processing appraisal information and Fraser and Lord

(1988) discovered that stimulus prototypicality directly affected perceptions and ratings. Prior research has also concluded that when evaluating managers at mid-level or higher, how closely the rated manager matches the rater's stereotype of a good manager, the higher the rating—holding performance results constant (Branson, Moe, & Sung, 2005). Branson Moe and Sung (2005) and Branson and Sung (2004) found that performance ratings conducted by both individuals and teams were significantly impacted by the stereotyping effect. The closer the rated manager matched the stereotype of a “good manager” held by the rater, the higher the rating—holding actual performance constant.

Kaplan and Norton (1992) made the claim that financial performance measures alone are inadequate for the task of effective performance evaluation. Lipe and Salterio (2000) found that supervisors only use part of the information contained in a BSC, but researchers in psychology as well as business and accounting, have continued to call for the use of more measures, more diversity in the measures, and more results/outcome measures in order to overcome the impact of subjective measures, like stereotyping (Kline & Sulsky, 2009; Moers, 2005).

METHODOLOGY

This paper reports on an empirical study into the efficacy of the BSC to overcome the stereotyping effect in performance ratings. This study is based on a course embedded case analysis conducted by 176 upper division and graduate students in accountancy and business courses. The average age of the participants was 30, with an average work experience of 11 years. Sixty-nine percent were female and thirty-one percent were male. As indicated by the demographics, all participants were adult learners with significant work place experience who were pursuing continuing professional development. All participants were familiar with the performance evaluation process, and the Balanced Scorecard.

Organizational context, social context, organizational/political processes and features of the personal relationships between rater and ratee can result in bias in the performance rating. The current study controls these variables by using a mixed-factorial quasi-experimental design implemented outside an actual organization where the influences of these processes are avoided. Participants were given a stimulus case where they were asked to assume the role of an upper level manager of eight subordinate mid level managers. They were asked to evaluate the performance of the eight subordinate managers based on the information in the case. The case was composed of three parts.

The first part was a behavioral description of 12 managerial behaviors of the 8 subordinate managers. The described behaviors were systematically manipulated between rated managers, and between cases. The stereotype fit measures the extent to which the supervisor/rater perceived the subordinate manager/leader to be performing their responsibilities in the way the supervisor thinks they should be performed. The stereotype fit score measures how closely the rated manager

fits the rater's stereotype of a good manager. Using a five point scale, participants were asked to rate their theoretical subordinates on the twelve behavioral items. They were asked to indicate if the behavior of the subordinate should display more or less of the target behavior than currently displayed. The five point response scale ranged from -2 (less than currently displayed,) to +2 (more than currently displayed). The stereotype fit score is the sum of squared scores. A low stereotype fit score indicates that the leader/manager is performing in a way that is consistent with the expectations of the rater. A high score indicates that the manager/leader is performing in a way that is inconsistent with the rater's prototype of a good leader/manager. The range of possible stereotype fit scores was from 0 to 48. The behaviors described in this study were based on the behaviors that differentiate between leaders/managers, a technique used since Mount and Thompson (1987). Categorization Theory maintains that participants will classify each manager/leader as a manager/leader or non-manager/leader based on how closely their behaviors match the prototype of a good leader/manager held by the observer/supervisor (Feldman, 1981). This categorization acts as a filter in future information searches and processing, and has proven to be a very robust factor in performance ratings (Branson, & Sung, 2004). This variable is a within-subjects repeated-measures variable.

Participants were also provided with outcome/results information in either the form of financial performance measures, or a set of 34 BSC measures, using the 4 categories of the BSC. The outcome measures were systematically manipulated according to the Latin Square research design so the influence of each variable could be evaluated. One hundred and ten participants were given outcome information in the form of financial performance indicators only. Sixty-six participants were given outcome information in the form of a more diverse, multi source balanced scorecard based feedback system.

The dependent variable was a performance rating for each of the eight managers being evaluated. This study had a forced choice dependent variable where the rating managers were told they had a budget of \$100,000 to allocate between the eight subordinate managers in the form of bonuses. They were asked to allocate the bonus for each manager as they thought appropriate based on all the information in the case which they thought relevant to the decision.

Categorization Theory maintains that once a person is categorized, future information about that person is filtered by the original categorization (Feldman, 1981). In fact, raters will give ratees high ratings on behaviors that were never observed, simply because such a behavior is assumed by the category into which the target person was categorized (Krzystofiak, Cardy, and Newman 1988). Consequently, lower performing individuals may receive higher performance appraisals than higher performing individuals, simply because they match certain characteristics or behaviors expected by a prototype guided rater. The question asked by this project is whether or not the BSC measures are more powerful than

just financial data in diminishing the impact of stereotyping on performance appraisals. Consequently our primary null hypothesis is:

H₀: There will be no difference in the influence of stereotyping on performance ratings when using BSC or financial outcome/results information.

Control variables in this study include age of rater (the age of the ratees are not identified in the case), gender of rater (gender of ratees is not identified in the case—ratees were identified only by a last name, e.g. Adams), and number of years of work experience of the rater. Finally, sometimes teams/groups can do a better job of making a decision based on the facts (results) than individuals (Branson, Steele, & Sung 2010a; Schulz-Hardt, Jochims, & Frey, 2002). Consequently, after individuals had made their rating decisions, they were placed into pre-assigned teams (of 3-4 persons) and asked to repeat the rating decision as a group. The ratings of the groups were examined to see if the group decision making process was able to further attenuate the impact of stereotyping on performance ratings. This resulted in a family of null hypotheses all of which expected to find no difference in the ratings due to age, gender, years of work experience, or individual versus group decision making.

RESULTS

Data were analyzed using SAS General Linear Model procedure. The analysis of the demographic data of age, gender, years of work experience resulted in the following:

1. For financial data only: age has no influence on bonus ($p = 0.34$); years of experience has no influence on bonus ($p = 0.35$); age has no influence on team bonus ($p = 1.00$); years of experience has no influence on team bonus ($p = 0.99$).
2. For BSC data: age has no influence on bonus ($p = 0.39$); years of experience has no influence on bonus ($p = 0.95$); age has no influence on team bonus ($p = 0.98$); years of work experience has no influence on team bonus ($p = 0.87$).
3. Gender was not a factor for bonus or team bonus for either individuals or teams. However, while there was no difference in the mean bonus awarded, it was found that the bonus awarded by females had more variance than bonuses awarded by males for both the individually awarded bonus ($p = 0.002$) and the team awarded bonus ($p = 0.0003$).

The control variables of age, gender, and years of work experience dropped out of the model. We continued our analysis by the use of SAS GLM procedure to identify the specific influence of stereotype information on the ratings given by individuals and groups when they had either financial results data or BSC results

data. We then used SAS to conduct a homogeneity of slopes test to determine if the influence of stereotype information was different when raters had financial results data only, or when they had BSC results data. We also tested the homogeneity of slopes for the ratings given by individuals and teams, both with financial results data, and with BSC results data.

Model 1.1 -- for individuals with stereotype and financial information:

$$\text{Bonus} = 17022 - 349.34 * \text{Stereotype Fit} + 3.45 * \text{Stereotype Fit}^2$$

$$R^2 = 0.15$$

Parameter	<i>B</i>	<i>t</i>	<i>p</i>
Intercept	17022	33.94	<.0001
Stereotype Fit	-349.34	-5.62	<.0001
Stereotype Fit ²	3.45	2.58	0.01

Table 1
Individuals with Stereotype Information and Financial Results

Model 1.2 – teams with stereotype and financial information:

$$\text{Team Bonus} = 16002 + 1950.91 * \text{Stereotype Fit}^{1/2} - 891.17 * \text{Stereotype Fit} + 9.36 * \text{Stereotype Fit}^2$$

$$R^2 = 0.25$$

Parameter	<i>B</i>	<i>t</i>	<i>p</i>
Intercept	16002	18.68	<.0001
Stereotype Fit ^{1/2}	1950.9	-4.86	<.0001
Stereotype Fit	-891.17	2.44	0.015
Stereotype Fit ²	9.36	4.68	<.0001

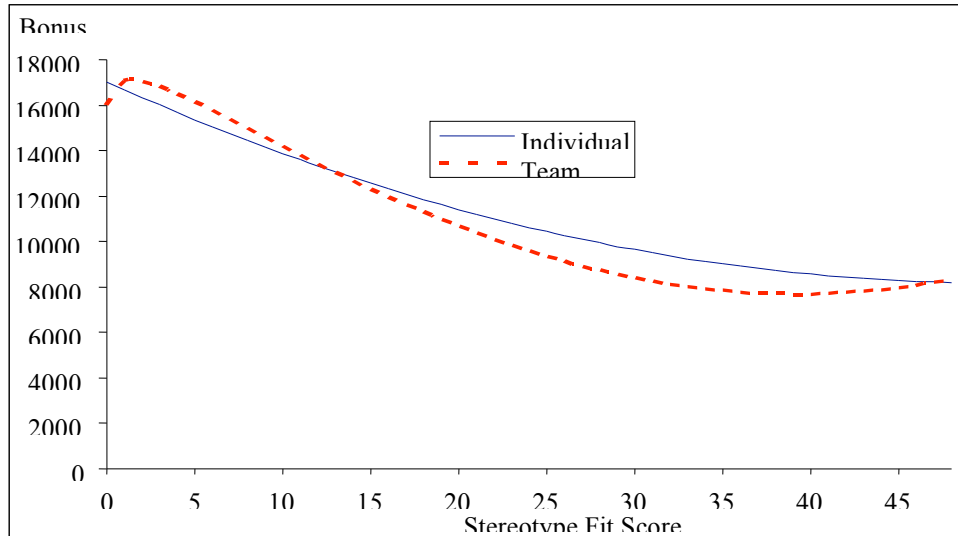
Table 2
Teams with Stereotype Information and Financial Results Data

Stereotype information has a significant impact on the bonus given by both individuals and teams when they have only financial results data. In the case of individuals with financial results data, 15% of the variance in the bonus (rating) is explained by the model based on how closely the rated manager fit the rater’s stereotype of a good manager. In the case of teams with financial results data, 25% of the variance in the rating was accounted for by the stereotype fit. How closely the rated manger fit the leader/manager stereotype of the rater had a significant impact on the rating, regardless of the results achieved by the rated manager.

Graph three and the homogeneity of slopes test indicated there was no significant difference in the slopes for individuals and teams with financial results data only.

Teams did not significantly reduce the influence of stereotype information on the ratings relative to individuals.

Graph 1
Stereotype Fit Score and Bonus/Rating-Financial Data
Individuals vs. Teams



When looking at the bonus allocation given by individuals and teams when they have stereotype fit and BSC information we found that stereotype fit information had a significant influence on the ratings for both individuals and teams.

Model 2.1 – individuals with Stereotype Fit and BSC information:

$$\text{Bonus} = 15133 - 135.27 * \text{Stereotype Fit}$$

$$R^2 = 0.047$$

Parameter	<i>B</i>	<i>t</i>	<i>p</i>
Intercept	15133	19.58	<.0001
Stereotype Fit	-135.27	-4.56	<.0001

Table 3
Individuals with Stereotype Fit Information and BSC Results Data

Model 2.2—teams with stereotype and BSC information

$$\text{Bonus} = 15788 - 159.26 * \text{Stereotype Fit}$$

$$R^2 = 0.08$$

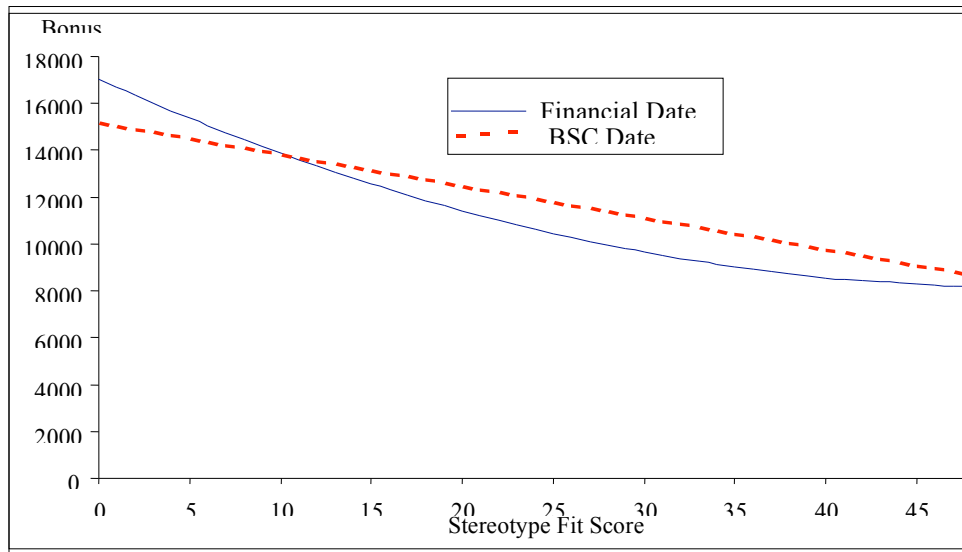
Parameter	<i>B</i>	<i>t</i>	<i>p</i>
Intercept	15788	22.92	<.0001
Stereotype Fit	-159.26	-6.02	<.0001

Table 4
Teams with Stereotype Fit Information and BSC Results Data

Stereotype fit information continued to have a significant impact on ratings given by both individuals and teams with BSC results information. When testing for homogeneity of slopes between the individuals and the teams with BSC data, there was no significant difference in the slopes of the regression functions. There was no significant difference in the impact of stereotype fit information for individuals and teams with BSC data.

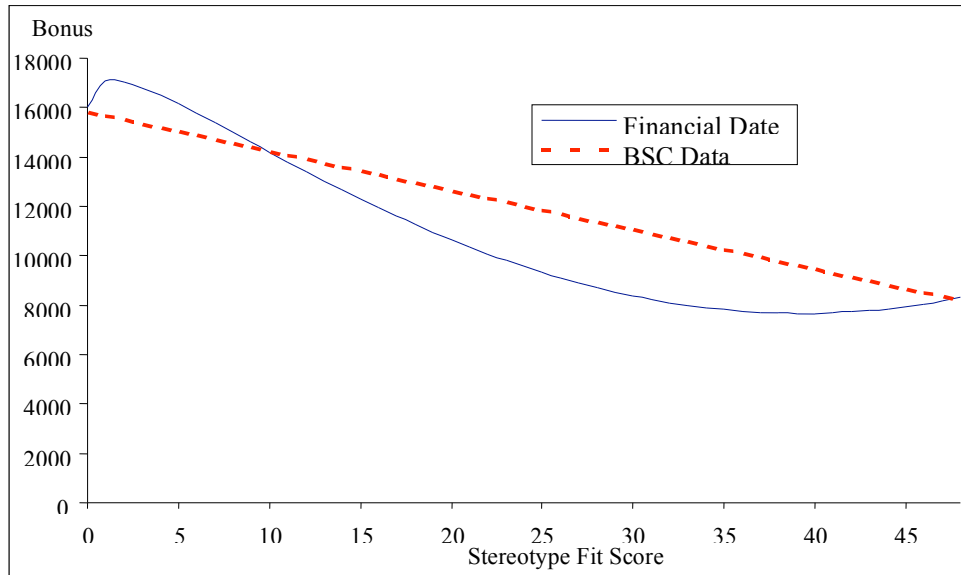
The central question in this research was whether or not the multiple source information provided by the BSC mitigates the influence of stereotype fit information on the ratings given by individuals and groups/teams, when compared to the use of financial outcome data alone. As indicated in Graph 2 when testing for the homogeneity of slopes between individuals using financial data and individuals using BSC data, there was a significant decrease ($p = 0.02$) in the influence of stereotype fit data when participants used BSC data.

Graph 2
Stereotype Fit Score and Bonus/Rating-Financial Data vs. BSC Data
Individuals



Graph 3 illustrates the slopes of the regression functions for team evaluations when using financial data and when using BSC data. Again SAS GLM was used to test for the homogeneity of slopes. In the case of teams with financial data and teams with BSC data, there was a significant decrease ($p = 0.002$) in the influence of stereotype fit on the performance ratings when BSC data was used.

Graph 3
Stereotype Fit Score and Bonus/Rating-Financial Data vs. BSC Data
Teams



CONCLUSIONS

Due to stress, time constraints, and heavy cognitive load placed on human decision makers we have developed cognitive tools to deal with the “overload”. We process information through a highly organized cognitive structure composed of cognitive categories into which we place all the information we gather. We use heuristics or mental shortcuts to process the never ending stream of information to which we are subjected. While these cognitive processes improve our efficiency in information processing, and allow us cognitive economy, they necessarily lead to the use of decision irrelevant and even made-up information in our decision making. Once the categories are “learned” they allow us to represent a highly complex external world in terms of a more abstract, but simpler cognitive representation. This representation serves as a guide for selective attention, for storage and processing of information, and for the formation of judgments.

When humans experience another person, we place them into a set of categories. These categories will direct future information search, future perceptions, and

future decision making about that person. Once a person is placed into a category, the person is recalled and thought about in terms of the prototype of the categories into which they have been placed, and not necessarily as their true selves. This process of stereotyping is omnipresent, and powerful in human interactions. The purpose of this study was to investigate the ability of a multiple source, diverse, multiple measure performance management system to mitigate the negative impact of stereotyping on performance ratings. Prior studies (Branson & Sung, 2004; Steele, 2010) found that stereotyping is a very robust phenomenon. This study looked specifically at the impact of stereotyping on performance ratings in the context of outcome information in the form of either financial results, or in the form of the Balanced Scorecard.

This study found that age, gender or years of work experience of the rater were not significant factors in the ratings. The age, gender and years of work experience of the ratees were not identified in the stimulus material, presumably eliminating their potential impact, and the impact of their interaction with rater demographics, on the dependent variable. Prior lab experiments and field studies have systematically found that non-performance factors have a major impact on performance ratings (e.g., Steele, 2010). This study is in line with that stream of research in that it was found that stereotyping, a cognitive process of the rater, did in fact impact the ratings, even in the presence of objective outcome/results information. This study extends the current body of knowledge in that it was found that the form in which the objective outcome information is provided can have a significant impact on how much influence stereotyping has on the ratings/dependent variable. The multiple cues, diverse information provided by the BSC do attenuate the impact of stereotyping on performance ratings in a lab experiment using professional adults who were not subject to the pressures of organizational life.

REFERENCES

- Bol, J. C., Smith, S. D. (2011). Spillover effects in subjective performance evaluation: Bias and the asymmetric influence of controllability. *The Accounting Review*, 86(4), 1213-1230.
- Branson, L., Sung, C.H. (2004). An empirical investigation of the role of accounting outcome information and group decision making in mitigating the affect of stereotyping on performance appraisal. *Proceedings of Academy of Business and Behavioral Sciences* annual conference, February 19-20, 2004.
- Branson, L., Moe, B., Sung, C. H. (2005). An empirical investigation of information processing in individuals, and virtual teams. *The Journal of Business and Behavioral Science*. 13(1), 234-242.
- Branson, L., Steele, N.L., Sung, C.H., (2010a). When two heads are better than one: Impact of group style and information type on performance evaluation. *Journal of Business and Behavioral Sciences*, 22(1), 75-84.

- Branson, L., Steele, N.L., Sung, C.H., (2010b). Differences in the decision making of constructively styled virtual and face-to-face teams regarding performance appraisal. *International Journal of Business and Public Administration*, 7(2), 11-22.
- Branson, L., Steele, N.L., Sung, C.H., (2011). Out of reach and out of touch: Differences in social awareness between and within collocated and virtual teams. Proceedings of Academy of Business and Behavioral Sciences, 18th Annual Conference, Las Vegas, February 18-19, 2011.
- Cardinaels, E., & van Veen-Dirks, P.M.G. (2010). Financial versus non-financial information: The impact of information organization and presentation in a balanced scorecard. *Accounting Organizations and Society*, 35, 565-578.
- Feldman, J.M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology*, 66, 127-148.
- Findley, Henry M., William F. Giles, and Kevin W. Mossholder. 2000. Performance appraisal process and system facets: Relationships with contextual performance. *Journal of Applied Psychology*, 85(4), 634-640.
- Greguras, Gary, Chet Robie, Deidra J. Schleicher, Maynard Goff III. 2003. A field study of the effects of rating purpose on the quality of multisource ratings. *Personnel Psychology*. 56, 1-21.
- Hastie, R., (1981). Schematic principles in human memory. *Social Cognition: The Ontario Symposium* ed. I. Higgins, T., Herman, C.P., Zanna, M.P., Hillsdale: Lawrence Erlbaum Associates Publishers, 39-88.
- Henneman, R., (1986). The relationship between supervisory ratings as results-oriented measures of performance: A meta-analysis. *Personnel Psychology*, 39, 811-827.
- Ilgen, D.R., & Feldman, J.M., (1983). Performance appraisal: a process focus, *Research in Organizational Behavior*, 5, 141-197.
- Ittner, C., Larcker, D.F., Meyer, M.W., (2003). Subjectivity and the weighting of Performance Measures: Evidence from a Balanced Scorecard. *The Accounting Review*. 78(3), 725-758.
- Ittner, C.D., Larcker, D.F., Randall, T., (2003). Performance implications of strategic performance measurement in financial services firms. *Accounting Organizations and Society*, 28, 715-741.
- Kaplan, R.S., & Norton, D.P., (1992). The balanced scorecard: Measures that drive performance, *Harvard Business Review*, 70(1), 71-79.
- Kim, T.Y., Cable, D., Kim, S.P., and Wang, J. (2009). Emotional competence and work performance: The mediating effect of proactivity and the moderating effect of job autonomy. *Journal of Organizational Behavior*. 30, 983-1000.
- Kline, T., Sulsky, L.M. (2009). Measurement and assessment issues in performance appraisal. *Canadian Psychology*, 3, 161-171.
- Krzystofiak, F., Cardy, R., Newman, J. (1988). Implicit personality and performance appraisal: The influence of trait inferences on evaluations of behavior. *Journal of Applied Psychology*. 73, 515-521.

- Landy, F.J., & Farr, J.L. (1980). Performance rating. *Psychology Bulletin*, 87(1), 72-107.
- Landy, F.J., & Farr, J.L. (1983). *The measurement of work performance methods: theory and applications*, New York: Academic Press.
- Lipe, M. & Salterio, S.E. (2000). The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review*, 75(3), 283-298.
- Moers, F. (2005). Discretion and bias in performance evaluation: The impact of diversity and subjectivity. *Accounting Organizations and Society*, 30, 67-80.
- Roberts, M.L., Albright, T.L., Hibbets, A.R. (2004). Debiasing balanced scorecard evaluations. *Behavioral Research in Accounting*, 16, 75-88.
- Schulz-Hardt, S., Jochims, M., & Frey, D. (2002). Productive conflict in group decision making: Genuine and contrived dissent as strategies to counteract biased information seeking. *Organizational Behavior and Human Decision Processes*, 88, 563-586.
- Spence, J.R., Keeping, L.M., (2010). The impact of non-performance information on ratings of job performance: A policy-capturing approach. *Journal of Organizational Behavior*, 31, 587-608.
- Steele, N. L. (2010, October). Differences in relevant information usage and group style for virtual and collocated decision making groups. Published in proceedings of fall meeting of the *International Academy of Business and Public Affairs Disciplines*, New Orleans, LA.
- Steele, N. L., Branson, L., Sung, C.H. (2011). Just how important are results? Rated importance and use of outcome information versus interpersonal cues in performance evaluation. *International Journal of the Academy of Business and Public Administration*. 8(1), 1-11.
- Taylor, S., Crocker, J., (1981). Schematic bases of social information processing. *In Social Cognition: The Ontario Symposium* ed. Higgins, E.T., Herman, C.P., Zanna, M.P., Hillsdale: Lawrence Erlbaum Associates Publishers, 39-88.
- van Veen-Dirks, P. (2010). Different uses of performance measures: The evaluation versus reward of production managers. *Accounting Organizations and Society*. 35, 141-164.
- Wang, Xiaoye M., Kin Fai Ellick Wong, and Jessica Y.Y. Kwong. 2010. The roles of rater goals and ratee performance levels in the distortion of performance ratings. *Journal of Applied Psychology*. 95(3), 546-561.
- Wendelken, D., & Inn, A., (1981). Nonperformance influences on performance evaluations: A laboratory phenomenon? *Journal of Applied Psychology*, 66, 149-158.
- Wong-On Wing, B., Gue, L., Wei, L., Yang, D. 2007. Reducing conflict in balanced scorecard evaluations. *Accounting Organizations and Society*, 32, 363-377.