BACK TO THE FUTURE: GENDER DIFFERENCES IN SELF-RATINGS OF TEAM PERFORMANCE CRITERIA

Christopher M. Scherpereel
Northern Arizona University
chris.scherpereel@nau.edu

Mary Y. Bowers
Northern Arizona University
mary.bowers@nau.edu

ABSTRACT

Well functioning teams are essential to success in both academic and businesses environments today. While much research has contributed to our understanding of successful team characteristics, conflicting empirical evidence hinders corrective intervention. Information gathered from self- and peer-assessments is often used to provide feedback and encourage teams to take corrective actions. Unfortunately, these assessments have been shown to contain many biases. In this study, we isolate gender as an important contributor to assessment bias. By exploring the gender difference in under- and overrating self-assessment relative to peer assessment, we add support to the assumption that gender plays a role in assessment accuracy and contribute new understanding to the differences in gender response to feedback.

INTRODUCTION

Teams are now considered an essential component of the workplace (Acona, 1990) and many organizations design jobs specifically for teams (Feder, Battenhausen, & Davis, 1999). As a result, a number of evaluative measures used to assess team performance and outcomes have been developed (Church & Bracken, 1997). Many of these evaluative measures have been used in the assessment of experiential exercises and simulation games. Gentry et.al. (2003) pointed out that while peer evaluations have become the standard method of assessing individual contribution to team performance, there are a number of issues regarding the use and structure of peer evaluations. While many of these issues have been discussed at ABSEL, one area that has received little attention is the impact of gender biases in peer- and self-ratings. Wheatley, Anthony, and Maddox (1988) recognized the importance of this issue by suggesting that "gender differences must certainly be pursued if we are to fully understand the efficacy of team participation in business simulations." (p. 136). We took this charge and empirically explored gender differences in under- and overrating self-assessment relative to peer assessment. In this study, we present support for the assumption that gender plays a role in assessment accuracy and also address the differences in gender response to feedback.

Because one of the primary missions of any business school is the preparation of students to function effectively in their careers, many business schools have incorporated teamwork and team assessment into the curriculum (Hansen, 2006). Teamwork in an academic setting generally involves projects, case studies, practice sets, computer simulations, and presentations. Overall, team assignments are considered by educators to provide a positive qualitative and quantitative learning experience for students. Team assignments promote the development of soft skills (written and oral communication, negotiation, conflict resolution) and the application of theory and concepts which promotes deeper understanding of discipline-specific topics. The value of incorporating teamwork in the classroom has been examined and confirmed in numerous studies (Price, 2004)

While the value of incorporating teamwork into business curriculum is well researched and documented, the validity of appraisal methods (particularly peer evaluations that include self-ratings, which are commonly used in the classroom) to assess team outcomes is not as clear or apparent (Cederblom & Lounsbury, 1980; Hansford & Hattie, 1982; Morahan-Martin, 1996). One of the most significant problems with peer evaluations is the observed discrepancies between self-ratings and peer ratings. It has been suggested by Franks, Ferguson, Rolls, & Henderson (1998) that successful performers are those whose self ratings are closer to peer ratings. More significantly, the accuracy of peer/self evaluations has been shown to have serious implications for self-confidence and psychological health (S. E. Taylor, Collins, Skokan, & Aspinwall, 1989).

LITERATURE REVIEW

Although, self/peer evaluations are widely used in both industry and academic settings, there is a lack of standardized measurement instruments and a number of conflicting theories on the optimal implementation (Ohland, Layton, Loughry, & Yuhasz, 2005). In addition, the problematic nature of self-assessment has long been

recognized (Atwater, Ostroff, Yammarino, & Fleenor, 1998) While some research reports students are accurate in their self-assessment (Brown, 1995), most studies strongly indicate that self-ratings are generally less accurate, more biased, and hence, less reliable than are peer ratings (Atwater & Yammarino, 1992; Fleenor, McCauley, & Brutus, 1996; Harris & Schaubroeck, 1988; Mabe & West, 1982; VanVelsor, Taylor, & Leslie, 1993). The inaccuracies are exposed by either inflated or deflated self-ratings. Different theories support these results.

HIGHER SELF-RATINGS

Much of the research conducted on self-rating concluded that self-ratings were often higher than peer ratings (Holzbach, 1978; Saavedra & Kwun, 1993; Thornton, 1980). In studies examining 360-degree feedback programs (an organizational self- and peer-rating system), self-ratings were found to be more tolerant and thus higher than peer ratings; this phenomenon is known as the leniency effect (VanderHeijden & Hijhof, 2004). The theoretical literature provides numerous explanations for higher self-ratings. These theories include the need for self-aggrandizement (S. Taylor & Brown, 1988); the need to maintain an ideal self-image (Harris & Schaubroeck, 1988); and the need to exaggerate self-ratings to enhance overall performance evaluation (Holzbach, 1978). In addition, higher self-ratings may be explained by applying attribution theory (Kelley, 1972): good performance is attributed to positive personal characteristics by self-raters while peer raters attribute the same performance to external factors (Devader, Bateson, & Lord, 1986).

These theories have been substantiated in a variety of research studies: poor performers tend to overestimate their abilities (Randall, Ferguson, & Patterson, 2000); inexperienced leaders provide inflated self assessments (Bass & Yammarino, 1991); and men tend to give inflated self-ratings (Beyer, 1992; Beyer & Bowden, 1997).

LOWER SELF-RATINGS

While the causes of inflated self-ratings highlighted above are fairly intuitive, explanations for underrating are somewhat more complex. Research indicates that the need to be self-effacing, self-assessment based on inaccurate or incomplete self-knowledge, use of an inappropriate comparison, or misunderstanding of the criteria being used for assessment are all possible reasons for self-underrating (Atwater, Ostroff, Yammarino, & Fleenor, 1998).

These negative self-perceptions could prevent individuals from pursuing opportunities for which they are qualified (resulting in underachievement), or could lead to greater individual effort in order to compensate (resulting in greater success). Beyer (1990) explored the accuracy of self evaluations and concluded that gender plays a significant role in the underrating of performance, with women consistently providing lower self-ratings.

PERSISTENCE IN SELF-RATINGS

Although the reasons for both a high and low self-rating bias are intuitively appealing, the one variable that seems to explain both rating errors is gender. Researchers have long considered gender to be a crude method for making empirical distinctions between groups (Deaux, 1984); however, it appears effective in explaining the differences in the observed self-rating bias. Previous studies have reported that males tend to overrate themselves, while females tend to underrate themselves across a variety of evaluation criteria. Nilsen and Campbell (1993) demonstrated that there is consistency over time in people's tendency to self overrate or underrate. The question is, whether these biased self-ratings are still persistent when individuals are provided with feedback on their performance. For example, when individuals are provided peer feedback on their performance do they continue to report biased self evaluations or do they adjust future ratings? Additionally, is there any difference between the consistencies of responses by males versus females (persistence)?

HYPOTHESES

These questions and issues lead to a number of testable hypotheses regarding self-rating bias and gender differences. Based on the reasoning for high and low self-rating bias discussed above, the following hypotheses were developed and tested in this study.

H1: Females tend to underrate their team performance (relative to their peer ratings) to a greater extent than males

    Corollary H1:   Males tend to overrate their team performance (relative to their peer ratings) to a greater extent than females.

H2: Females continue to underrate their team performance (relative to their peer ratings) after receiving peer feedback

    Corollary H2:   Males continue to overrate their team performance (relative to their peer ratings) after receiving peer feedback

SETTING

DESCRIPTION OF BIZBLOCK

This paper focuses on students' self-ratings versus peer ratings in a junior level business course that is part of the curriculum at Northern Arizona University's W.A. Franke College of Business. This course, known as BizBlock, combines management, marketing, and business communication and gives students the opportunity to integrate key business concepts. Student teams create and

present a comprehensive business plan to a panel of real business leaders and learn to manage group performance.

The BizBlock design mandate was simple in theory: take the core three-credit-hour undergraduate courses

TABLE 1
Behavioral Scales

Communication	Listens attentively to others without interrupting
	Conveys interest in what others are saying
	Provides others with constructive feedback
	Restates what has been said to show understanding
	Clarifies what others have said to ensure understanding
	Articulates ideas clearly and concisely
	Uses facts to get points across to others
	Persuades others to adopt a particular point of view
	Gives compelling reasons for ideas
	Wins support from others
Decision Making	Analyzes problems from different points of view
	Anticipates problems and develops contingency plans
	Recognizes the interrelationships among problems and issues
	Reviews solutions from opposing perspectives
	Applies logic in solving problems
	Plays devil's advocate role when needed
	Challenges the way things are being done
	Solicits new ideas from others
	Generates new ideas
	Accepts change
	Suggests new approaches to solving problems
	Offers solutions based on facts rather than gut feel or intuition
	Discourages others from rushing to conclusions without facts
	Organizes information into meaningful categories
	Helps others to draw conclusions from the facts
	Brings in information from outside sources to help make decisions
Collaboration	Acknowledges issues that the team needs to confront and resolve
	Encourages ideas and opinions even when they differ from his/her own
	Works toward solution and compromises that are acceptable to all
	Helps reconcile differences of opinion
	Accepts criticism openly and no-defensively
	Shares credit for success with others
	Cooperates with others
	Encourages participation among all participants
	Shares information with others
	Reinforces the contributions of others
Self-Management	Monitors progress to ensure that goals are met
	Creates action plans and timetables for work session goals
	Defines task priorities for work sessions
	Ensures that goals are understood by all
	Puts top priority on getting results
	Stays focused on the task during meetings
	Uses meeting time efficiently
	Suggests ways to proceed during work sessions
	Clarifies roles and responsibilities of others
	Reviews progress throughout work sessions
	Solicits input from all members
	Encourages frequent polling about team members
	Summarizes the teams' position on issues
	Involves others in decisions that affect them

(which must be completed to earn a degree in business administration or accountancy) in management, marketing, and business communications and integrate the material so it can be delivered in a single nine-credit-hour course block. The key directive of this mandate was to integrate the three core courses, not just deliver the content of the three courses sequentially. To further reinforce the concept of integration, most of the work is done in teams that are formed the second day of class.

As such, the primary outcome assessment in BizBlock is a business plan created and developed by student teams. The students are organized in teams consisting of five to seven members, depending on the class size. The student teams are presented the problem of identifying a consumer need and then developing a business plan that fills that need. Lectures, assignments, exams, and activities are designed to motivate students to enhance, improve, and augment their understanding of the business. The resulting business plan is developed and revised throughout the semester-long course. Final plans are presented in a competitive format before a panel of 3-5 venture capitalists that provides outside validation to the students' work.

The final business plan and presentation makes up more than 50% of the grade in each class (marketing, management, and business communication) but receives a team grade. However, an individual student may receive a lower grade than their teammates on the plan and presentation as a result of a peer evaluation process that is an integral part of the course. Teams complete peer evaluations three times throughout the semester. The results of the first two evaluations are distributed to each team. As such, team members see how they were rated by the peers versus how they rated themselves. Feedback from the early evaluations is to be used to address and correct any member behaviors that negatively affect the performance of the team. If the last peer evaluation (done soon after the final plan is submitted) indicates a team member did not change dysfunctional behaviors indicated on earlier evaluations and did not contribute to the team's efforts, the team member's grade will be adjusted down appropriately. Therefore, failure to use developmental feedback effectively can result in a lower grade for the course.

DESCRIPTION OF RATING CRITERIA

A standard peer/self evaluation instrument called *The Team Developer* was used to collect data in this study. *The Team Developer* is a computer based survey developed by McGourty and DeMeuse (2001) consisting of 50 questions that measure four behavioral dimensions: Communication, Decision Making, Collaboration, and Self-Management. Each dimension is defined by several behavioral scales that indicate effectiveness as indicated in Table 1.

Team members use a five point rating scale to anonymously rate self and team members on various team behaviors. Based on the reported ratings, each team member is given a feedback report that pairs self-ratings with team members' (peer) ratings.

FINDINGS

Historically, a BizBlock class has more male students than female students and has more students enrolled in the fall than in the spring. These statistics have been consistent

TABLE 2
BizBlock Class Configuration

	GPA	Fall 2006		Spring 2007		Overall
		Number	Percent	Number	Percent	Percent
Females	3.12	27	42%	9	30%	38%
Males	3.11	37	58%	21	78%	62%
Total		64		40		104

TABLE 3
Proportion of Students Underrating Self

		Communication	Decision Making	Collaboration	Self Management
1st Peer Evaluation	Males	48.1%	44.4%	42.6%	48.1%
	Females	66.7%	60.6%	69.7%	45.5%
	Probability	0.046	0.072	0.007	0.404
2nd Peer Evaluation	Males	41.2%	45.1%	41.2%	49.0%
	Females	58.6%	51.7%	58.6%	51.7%
	Probability	0.067	0.284	0.067	0.408
3rd Peer Evaluation	Males	32.0%	30.0%	28.0%	30.0%
	Females	44.4%	51.9%	51.9%	51.9%
	Probability	0.139	0.030	0.019	0.030

since the class was first offered in fall 2000. The BizBlock class/study configuration for 2006-2007 is shown in Table 2.

It is interesting to note that the average GPA for both males and females was essentially the same; thus, the relationship studies by Randall, Ferguson, and Patterson (2000) indicating that high performers tended to underrate self and low performers tended to overrate self was controlled for in this study.

We specifically examine data collected from peer evaluations administered in BizBlock during the academic year 2006-2007. This data was gathered from the *Team Developer* peer evaluation format that had not been used in prior years.

Table 3 presents the proportion of BizBlock students who provided self-ratings lower than their peer ratings. Testing for the significance of the difference between two independent proportions provides the one tail probability that the two proportions are the same. For example, testing H1 {Females tend to underrate their team performance (relative to their peer ratings) to a greater extent than males} the probability is only 4.6% (in Table 3) that females and males underrate themselves on communication equally. Given the proportion data collected for the first peer evaluation, it can be concluded that females underrate themselves to a greater extent than males on the dimensions of communication, decision making, and collaborations (with no significant difference indicated on the self-management dimension).

Similar results were obtained in the second evaluation administered four weeks later. However, only the communication and the collaboration dimensions show a significant difference. And finally, Table 3 shows the results for the final evaluation (3rd peer evaluation), administered two weeks prior to the end of the semester. Again, females tended to underrate themselves to a greater extent than males, with the dimensions of decision making, collaboration, and self-management showing a significant difference.

Parallel results are observed in Table 4 where male students tend to overrate themselves when compared with female students {or the testing of Corollary H1: Males tend to overrate their team performance (relative to their peer ratings) to a greater extent than females}. It is interesting to note that this overrating tendency by males actually increases as students are provided peer feedback from previous evaluations. This observation appears consistent with the research of Taggar and Neubert (2004) that reports the indignation felt after receiving lower than expected peer ratings may result in anger and an upward bias in future self rating. Although this tendency is not noted to be gender specific in prior research, the data presented in Table 4 suggests females tend to adjust self-ratings to be consistent with peer ratings and males tend to adjust their self-ratings counter to peer ratings. This difference is shown to be significant when we look at the data for males and females separately.

Do female students consistently underrate themselves? This is the question that motivated the second hypothesis (H2): Females continue to underrate their team performance (relative to their peer ratings) after receiving peer feedback. Table 5 shows that in the initial evaluation, females signifcantly underrate themselves relative to their peers on the dimensions of communication, decision making, and collaboration. However, this underrating tendency seems to disappear in successive evaluations. In the final evaluation, it appears women and men are equally likely to overrate as well as underrate themselves. This outcome is clearly an expected response to receiving feedback and responding to this feedback by adjusting subsequent self evaluations. It is consistent with the research which indicates that accuracy of performance predictions increases with successive evaluation. (Radhakrishnan, Arrow, & Sniezek, 1996; Shepperd, Ouellette, & Fernandez, 1996) Specifically, Roberts and Holen-Hoeksema (1989) noted that women modify future self-evaluations in response to feedback from others.

Do male students consistantly overrate themselves? This is the question that motivated the corollary to the second hypothesis {Corollary H2: Males continue to overrate their team performance (relative to their peer ratings) after receiving peer feedback}. Table 6 shows that

TABLE 4
Proportion of Students Overrating Self

		Communication	Decision Making	Collaboration	Self Management
1st Peer Evaluation	Males	51.9%	51.9%	57.4%	50.0%
	Females	33.3%	39.4%	30.3%	54.5%
	Probability	0.046	0.129	0.007	0.340
2nd Peer Evaluation	Males	58.8%	52.9%	58.8%	51.0%
	Females	41.4%	48.3%	41.4%	48.3%
	Probability	0.067	0.344	0.067	0.408
3rd Peer Evaluation	Males	66.0%	70.0%	72.0%	68.0%
	Females	48.1%	48.1%	48.1%	48.1%
	Probability	0.064	0.030	0.019	0.044

TABLE 5
Proportion of Female Students Under/Over-rating Self

		Communication	Decision Making	Collaboration	Self Management
1st Peer Evaluation	Underrating	66.7%	60.6%	69.7%	45.5%
	Overrating	33.3%	39.4%	30.3%	54.5%
	Probability	0.003	0.042	0.001	0.230
2nd Peer Evaluation	Underrating	58.6%	51.7%	58.6%	51.7%
	Overrating	41.4%	48.3%	41.4%	48.3%
	Probability	0.095	0.396	0.095	0.396
3rd Peer Evaluation	Underrating	44.4%	51.9%	51.9%	51.9%
	Overrating	48.1%	48.1%	48.1%	48.1%
	Probability	0.392	0.393	0.393	0.393

TABLE 6
Proportion of Male Students Under/Over-rating Self

		Communication	Decision Making	Collaboration	Self Management
1st Peer Evaluation	Underrating	48.1%	44.4%	42.6%	48.1%
	Overrating	51.9%	51.9%	57.4%	50.0%
	Probability	0.350	0.221	0.062	0.439
2nd Peer Evaluation	Underrating	41.2%	45.1%	41.2%	49.0%
	Overrating	58.8%	52.9%	58.8%	51.0%
	Probability	0.037	0.214	0.037	0.422
3rd Peer Evaluation	Underrating	32.0%	30.0%	28.0%	30.0%
	Overrating	66.0%	70.0%	72.0%	68.0%
	Probability	0.000	0.000	0.000	0.000

in the initial evaluation males tend to equally underrate as well as overrate themselves although there is a significant tendency to overrate self in the area of collaboration. Interestingly, after receiving feedback, the data shows males tend to overrate themselves by a greater extent in subsequent evaluations, showing a signficant difference on all four dimensions by the final evalaution. The difference between male and female adjustment, when aggregated, may explain why some studies have concluded there is no improvement in evaluation accuracy across time. (Gordon, 1991; Powel & Gray, 1995) Do males not believe the feedback they are getting? The appearance of truth in this statement may be due to men seeing peer evlauations as inaccurate indicators of actual performance. The research indicates that overrating individuals tend to treat others' evaluations with skepticism. (Roberts & Nolen-Hoeksema, 1989)

Research conducted in the 1970s reported women were more inclined to rate themselves lower than men across a wide spectrum of performance related criteria. Retrospectively, this phenomenon could be explained by socially indoctrinated norms that implicitly required women to be modest and self-effacing. At this point in time, these norms were still pervasive in American culture despite the growing support and acceptance of the feminist movement.

In their extensive literature review of gender-based differences in attitudes and behavior, Maccoby and Jacklin (1974) found women consistently reported lower self-evaluations of abilities and expectancies. In addition, the literature also indicated women had fewer positive self-referent attitudes than men. The conclusions regarding women's proclivity for underrating put forward in this review were corroborated by studies done just a few years later (Deaux, 1979; Deaux & Farris, 1977; Hanlan, 1977).

However, by the late 1970s and early 1980s, research results indicating there were no gender differences in self-ratings began to appear. Snyder and Bruling (1979) concluded that women were no less likely than men to exhibit positive self-referent, job-related attitudes. Two studies done in 1981 reported that when ratings are kept confidential, there was no difference between male and female self-ratings (Berg & Dodson, 1981; Eagly & Carli, 1981). In addition, a study of male and female managers found that women's self-ratings were actually higher than those of their male counterparts (Tsui & Gutek, 1984). Perhaps the change in self-ratings by women from 1970s to 1980 reflected the change in attitudes towards women's equality (at least in terms of performance evaluations and self-ratings) because little research on gender differences in self-ratings was conducted during the rest of the 1980s and

early 1990s. In 1998, a study conducted with 91 senior marketing students found the difference between mean self-ratings by males and females was not statistically significant. Interestingly, the findings did indicate that "in today's university classrooms, both males and females appear to overestimate their own performances" (Haas, Hass, & Wotruba, 1998).

BACK TO THE FUTURE

Now, in 2007, a study of junior level business students replicates the results of self-rating studies conducted thirty years ago in which self-ratings by women were significantly lower than men.

One possible explanation for these results might be found in the literature regarding the impact of feedback on men's and women's self-ratings. Roberts (1991) found women were more likely to assimilate and apply information received from feedback than are men. Roberts and Nolen-Hoeksema (1989) found that self-rating by women was more likely to be modified to reflect feedback they receive from others than are men's self-ratings. In other words, women will change their self-rating to mirror peer ratings but men often will not because they do not believe peer ratings are as accurate as their self-ratings. Based on their earlier study, Roberts and Nolen-Hoeksema (1994) developed and tested three hypotheses regarding women's tendency to try to diminish differences between self- and peer ratings by adjusting their self-ratings. A hypothesis suggesting women and men don't perceive positive and negative feedback the same was not supported nor was a hypothesis that women underrated in order to adhere to the accepted feminine behavior of being modest. The hypothesis that women believed peer ratings to be more accurate and adjusted their self-ratings accordingly was supported. The lower self-rating by women presented in this paper may be reflective of the extensive feedback that is an integral part of BizBlock.

Another possible explanation of these results maybe that the women students evaluated their teamwork abilities using different criteria or more demanding standards than did their male counterparts. With this explanation, lower self-ratings would not be reflective of perceived lower abilities and contribution.

A third possible explanation may be what some researchers have identified as the importance of context in creating, erasing, or even reversing gender difference. (Hyde, 2005) In this study, there was a predominately male composition of both classes. The possibility that women in these two classes were intimidated enough by the predominance of males to impact their self-ratings while improbable is not impossible. This theory has some merit given that in the class with the largest proportion of males, most of the women rated themselves lower than their peer ratings in two of the rating criteria.

IMPLICATIONS

Recognizing that there remains a gender difference in self-evaluation has some significant implications for both practitioners and educators. Research has shown that the accuracy of self-evaluations is critical to career success. Beyer (1990) noted that low self assessments might lead to low expectancies and decreased performance; thus, persistent inaccuracies in self-assessment may be self-fulfilling. Empirical data seems to support the link between performance and self-assessment accuracy. In Fletcher and Kerslake's (1992) study of job candidates' self-evaluations and subsequent performance found inaccurate self raters were more likely to perform poorly on the job. McCall and Lombardo (1990) found accurate self-evaluation was positively related to the likelihood of being promoted, and they later (1993) found that overrating self was associated with career derailment. The relationship between accurate self-rating and successful performance was also noted by Franks et.al. (1998) and Atwater and Yamarino (1997).

Reduced performance is not the result of biased self-ratings. Ashford's (1989) review of the literature identifies a connection between accurate self-evaluation with job satisfaction and positive attitudes toward supervisors. In addition, overraters may exhibit hostility and resentment (Yammarino & Atwater, 1993). Randall, Ferguson, and Patterson (2000) noted a persistent overrating of self may not be conducive to learning from mistakes. Similar problems exist for underraters, where studies have shown both the positive benefit of surprise (receiving responses that are better than expected) and the negative effect of a lower overall sense of self-worth (Yammarino & Atwater, 1993). Research on self assessments among accurate self-raters have identified the common traits associated with effective leadership such as self-esteem, confidence, intelligence, and internal locus of control (VanVelsor, Taylor, & Leslie, 1993).

Thus, identification of persistent over- and underrating is critical to developing programs to improve performance, learning, and leadership. If there exists a gender difference in the tendency to under- or overrate self (as this study shows), the implications are profound. Educational programs should be developed to encourage accurate self-rating (Atwater & Yammarino, 1997). Feedback should be tailored to fit the rater tendency (Roberts & Nolen-Hoeksema, 1989). Self-rating should be cautiously implemented in 360 degree feedback programs (Fletcher & Kerslake, 1992). And, the use of self-ratings as a method of job selection (Randall, Ferguson, & Patterson, 2000) should be used with sensitivity to these findings.

FUTURE RESEARCH

What factors lead to student self- rating biases? Although this study looked exclusively at gender differences as influencing self-ratings, previous research has suggested numerous other factors play a role in under- and overrating biases. Boud and Falchikov's (1989) review of some of the early literature suggests self-rating biases appear due to

differences in student abilities, experience level of respondents, purpose of the assessment (developmental versus evaluative), participant characteristics (age, gender, ethnicity), subject matter, and the types of rating scales used. Future investigation designed to isolate and explore the interaction of these factors might explain the differences in empirical results and lead to a more accurate, descriptive model of self-assessment.

The question of improvement in self-rating accuracy improvement with practice has been explored in the literature with mixed results. Studies by Murstein (1965), Larson (1978), and McGeever (1978) indicated no change in self-rating accuracy over time while Peterson (1979), Arnold et. al. (1985) and Cowan (1988) showed a significant improvement in self-rating accuracy with practice. Given the mixed results and the data collected in this study, it appears future studies that include gender differences may provide an empirical explanation for differences in self- rating accuracy improvement over time.

In an article questioning whether agreement between self-rating versus peer rating matters, Atwater et.al. (1998) conclude that the relationship between self- and peer-ratings is complex and affects team interpersonal relationships to even a greater extent than it does team productivity. They emphasize the importance of considering the magnitude and direction of disagreement between self- and peer ratings in future research efforts. Given the known psychological impact of consistently underrating self (relative to others) on the rater's self-confidence (S. E. Taylor, Collins, Skokan, & Aspinwall, 1989), it is important that future research consider the results of this study, which show a consistent gender bias that affects the accuracy of self-evaluations.

REFERENCES

Acona, D. G. (1990). Outward Bound: Strategies for team survival in an organization. *Academy of Management Journal, 33*(June), 334-365.

Arnold, L., Willoughby, T. L., & Calkins, E. V. (1985). Self-evaluation in undergraduate medical education: a longitudinal perspective. *Journal of Medical Education, 60*, 21-28.

Ashford, S. J. (1989). Self-assessments in organizations: A literature review and integrative model. *Research in Organizational Behavior, 11*, 133-174.

Atwater, L. E., Ostroff, C., Yammarino, F. J., & Fleenor, J. W. (1998). Self-other agreement: Does it really matter? *Personnel Psychology, 51*, 577-598.

Atwater, L. E., & Yammarino, F. J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions. *Personnel Psychology, 45*, 141-164.

Atwater, L. E., & Yammarino, F. J. (1997). Self-other rating agreement: A review and model. *Research in Personnel and Human Resource Management, 15*, 121-174.

Bass, B. M., & Yammarino, F. J. (1991). Congruence of self and other's leadership ratings of naval officers for understanding successful performance. *Applied Psychology: An International Review, 40*, 437-454.

Berg, S. J., & Dodson, M. (1981). Attributional modesty in women. *Psychology of Women Quarterly, 5*, 711-727.

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology, 59*(5), 960-970.

Beyer, S. (1992). *Self-consistency and gender differences in the accuracy of self evaluations.* Paper presented at the Annual Convention of the American Psychological Association, Washington, DC, USA.

Beyer, S., & Bowden, E. M. (1997). Gender differences in self perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin, 23*(2), 157-172.

Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: a critical analysis of findings. *Higher Education, 18*, 529-549.

Brown, R. W. (1995). *Autorating: getting individual marks from team marks and enhancing teamwork.* Paper presented at the 1995 Frontiers in Education Conference, Atlanta, GA, USA.

Cederblom, D., & Lounsbury, J. W. (1980). An investigation of user acceptance of peer evaluations. *Personnel Psychology, 33*(3), 567-579.

Church, A. H., & Bracken, D. W. (1997). Advancing the state of the art of 360-degree feedback. *Group & Organizational Management, 22*(2), 149-161.

Cowan, J. (1988). Struggling with self-assessment. In D. J. Boud (Ed.), *Developing Student Autonomy in Learning* (2 ed., pp. 192-210). London: Kogan Page.

Deaux, K. (1979). Self-evaluation of male and female managers. *Journal of Sex Roles, 5*(5), 571-580.

Deaux, K. (1984). From individual differences to social categories: Analysis of a decade's research on gender. *American Psychologist, 39*, 105-116.

Deaux, K., & Farris, E. (1977). Attributing causes for one's own performance: The effects of sex, norms, and outcome. *Journal of Research in Personality, 11*(March), 59-72.

Devader, C. L., Bateson, A. G., & Lord, R. G. (1986). Attribution theory: A meta-analysis of attribution hypothesis. In E. Locke (Ed.), *Generalizing from laboratory to field settings* (pp. 63-81). Lexington, MA: Lexington Books.

Eagly, A., & Carli, L. (1981). Sex of researchers and sex-typed communications as determinants of sex differences influenceability: A meta-analysis of social influence studies. *Psychological Bulletin, 90*, 1-20.

Feder, D. B., Battenhausen, K. L., & Davis, W. (1999). Peer reviews: Employees dual roles as raters and recipients. *Group & Organizational Management, 24*, 92-120.

Fleenor, J. W., McCauley, C., & Brutus, S. (1996). Self-other rating agreement and leader effectiveness. *Leadership Quarterly, 7*, 487-506.

Fletcher, C., & Kerslake, C. (1992). The impact of assessment centres and their outcomes on participants' self-assessments. *Human Relations, 45*, 73-81.

Franks, D., Ferguson, E., Rolls, S., & Henderson, F. (1998). Self-assessments in HRM: An example from an assessment centre. *Personnel Review, 28*, 124-133.

Gentry, J., Howard, B., Vaughan, M. J., Cudworth, A., & Vik, G. (2003). Improving the effectiveness of peer evaluations. *Developments in Business Simulation and Experiential Learning, 30*, 109-112.

Gordon, M. J. (1991). A review of the validity and accuracy of self-assessments in health profession training. *Academic Medicine, 66*, 762-769.

Haas, A. L., Hass, R. W., & Wotruba, T. R. (1998). The use of self-ratings and peer ratings to evaluate performance of student group members. *Journal of Marketing Education, 20*(3), 200-209.

Hanlan, M. S. (1977). Women in social work administration: Current role strains. *Administration in Social Work, 1*, 259-265.

Hansen, R. S. (2006). Benefits and problems with student teams: Suggestions for improving team projects. *Journal of Education for Business, 82*(1), 11-19.

Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research, 52*(1), 123-142.

Harris, M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology, 41*, 43-62.

Holzbach, R. L. (1978). Rater bias in performance ratings: Supervisor, self, and peer ratings. *Journal of Applied Psychology, 63* (October), 579-588.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*(6), 581-592.

Kelley, H. H. (1972). Attribution in Social Interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins & B. Weiner (Eds.), *Attributions: Perceiving the Causes of Behavior*. Morristown, NJ: General Learning Press.

Larson, M. B. (1978). Multiple copies of exams encourage self-grading. *Engineering Education, 68*(5), 435-437.

Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: a review and meta-analysis. *Journal of Applied Psychology, 67*, 280-290.

Maccoby, E. E., & Jacklin, C. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

McGeever, P. J. (1978). Student self-grading in the introductory American politics course. *Teaching Political Science, 5*(3), 319-330.

McGourty, J., & DeMeuse, K. P. (2001). *The Team Developer: An assessment and skill building program*. Hoboken: John Wiley & Sons.

Morahan-Martin, J. (1996). Should peers' evaluations be used in class projects? Questions regarding reliability, leniency, and acceptance. *Psychological Reports, 78*(3), 1243-1251.

Murstein, B. I. (1965). The relationship of grade expectations and grades believed to be deserved to actual grades received. *Journal of Experimental Education, 33*(4), 357-362.

Nilsen, D., & Campbell, D. P. (1993). Self-observer rating discrepancies: Once an overrater, always and overrater? *Human Resource Management, 32*, 265-281.

Ohland, M. W., Layton, R. A., Loughry, M. L., & Yuhasz, A. G. (2005). Effects of behavioral anchors on peer evaluation reliability. *Journal of Engineering Education, 94*(3), 319-326.

Peterson, J. M. (1979). Me and my critics: students' responses to architectural jury criticism. *Studies in Art Education, 20*(2), 64-67.

Powel, W. D., & Gray, R. (1995). Improving performance predictions by collaboration with peers and rewarding accuracy. *Child Study Journal, 25*, 141-154.

Price, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education, 93*(3), 223-231.

Radhakrishnan, P., Arrow, H., & Sniezek, J. A. (1996). Hoping, performing, learning, and predicting: Changes in the accuracy of self-evaluations of performance. *Human Performance, 9*, 23-49.

Randall, R., Ferguson, E., & Patterson, F. (2000). Self-assessment accuracy and assessment centre decisions. *Journal of Occupational and Organizational Psychology, 73*, 443-459.

Roberts, T.-A. (1991). Gender and the influence of evaluations on self-assessments in achievement settings. *Psychological Bulletin, 109*, 297-308.

Roberts, T.-A., & Nolen-Hoeksema, S. (1989). Sex differences in reactions to evaluative feedback. *Sex Roles, 21*(11/12), 725-746.

Roberts, T.-A., & Nolen-Hoeksema, S. (1994). Gender comparisons in responsiveness to others' evaluations in achievement settings. *Psychology of Women Quarterly, 18*(2), 221-240.

Saavedra, R., & Kwun, S. K. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology, 78* (June), 450-462.

Shepperd, J. A., Ouellette, J. A., & Fernandez, J. K. (1996). Abandoning unrealistic optimism: Performance estimates and the temporal proximity of self-relevant feedback. *Journal of Personality and Social Psychology, 70*, 844-855.

Snyder, R. A., & Bruling, N. S. (1979). Sex differences in perceived competence: An across organizations study. *Administration in Social Work, 3*(March), 349-358.

Taggar, S., & Neubert, M. (2004). The impact of poor performances on team outcomes: An empirical examination of attribution theory. *Personnel Psychology, 57*, 935-968.

Taylor, S., & Brown, J. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin, 103*, 193-210.

Taylor, S. E., Collins, R. L., Skokan, L. A., & Aspinwall, L. G. (1989). Maintaining positive illusions in the face of negative information: Getting the facts without letting them get you. *Journal of Social and Clinical Psychology, 8*(2), 114-129.

Thornton, G. C. (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology, 33*(summer), 263-271.

Tsui, A., & Gutek, B. (1984). A role set analysis of gender differences in performance, affective relationships and career success of middle managers. *Academy of Management Journal, 27*, 619-635.

VanderHeijden, B. I. J. M., & Hijhof, A. H. J. (2004). The value of subjectivity: problems and prospects for 360-degree appraisal systems. *International Journal of Human Resource Management, 15*(3), 493-511.

VanVelsor, E., Taylor, S., & Leslie, J. B. (1993). An examination of relationships among self-perception accuracy, self-awareness, gender and leader effectiveness. *Human Resource Management, 32*(2/3), 249-263.

Wheatley, W. J., Anthony, W. P., & Maddox, E. N. (1988). The relationship of locus of control and vividness of imagination measures to simulation performance. *Developments in Business Simulation and Experiential Learning, 15*, 134-137.

Yammarino, F. J., & Atwater, L. E. (1993). Understanding self-perception accuracy: Implications for human resource management. *Human Resource Management, 32*(2/3), 231-247.