


2017

Participant Self-Assessment of Development Center Performance

Ryan Powley

Minnesota State University, Mankato

Follow this and additional works at: <http://cornerstone.lib.mnsu.edu/etds>

 Part of the [Cognition and Perception Commons](#), [Human Resources Management Commons](#),
and the [Industrial and Organizational Psychology Commons](#)

Recommended Citation

Powley, Ryan, "Participant Self-Assessment of Development Center Performance" (2017). *All Theses, Dissertations, and Other Capstone Projects*. 704.

<http://cornerstone.lib.mnsu.edu/etds/704>

This Thesis is brought to you for free and open access by the Theses, Dissertations, and Other Capstone Projects at Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato. It has been accepted for inclusion in All Theses, Dissertations, and Other Capstone Projects by an authorized administrator of Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato.

Participant Self-Assessment of Development Center Performance

Ryan W. Powley

A Thesis Submitted in Partial Fulfillment for the Degree of Master of Arts in
Industrial/Organizational Psychology

Minnesota State University, Mankato

Mankato, Minnesota

April 7, 2017

Participant Self-Assessment of Development Center Performance

This thesis has been examined and approved by the following members of the student's committee.

Dr. Andrea Lassiter, Advisor

Dr. Daniel Sachau, Committee Member

Dr. Chris Brown Mahoney, Committee Member

Abstract

In order to navigate the political and cultural realities of modern organizations, individuals must possess an accurate self-perception (Ostroff, Atwater, & Feinberg, 2004). One way to gauge the accuracy of a person's self-perception is to compare that individual's self-assessment against other people's assessments of him or her. This technique is known as self-other rating agreement (SORA). Heneman (1980) posited that development centers are an ideal setting in which to study self-assessment and SORA. The present study examined self-assessment and SORA in development centers. Particular attention was paid to exercises, competencies, and gender differences in self-ratings and SORA. Correlations and t-tests were conducted to investigate gender effects, self-rating tendencies, and SORA of participants' self-assessments of performance. Results revealed that men self-rated higher than observers on 4 of 6 exercises and 4 of 7 competencies, women self-rated accurately on 4 of 6 exercises and 6 of 7 competencies, and men self-rated higher than women on all competencies and all but one exercises. This study's sample size was undesirably small, which unfortunately precluded the investigation of two proposed hypotheses. While these findings do little to advance the theory behind SORA, they still contribute to existing literature about gender, self-assessment, and development center exercises and competencies. Provided a much larger sample size could be obtained, future research should further investigate gender effects, self-rating tendencies, and SORA of participants' self-assessments of development center performance, in the hopes of helping participants improve their self-perception.

Keywords: self-assessment, development centers, self-other rating agreement

Table of Contents

Background	5
Development Centers	6
Self-Other Rating Agreement	6
Factors Affecting SORA	7
SORA in Centers	8
Present Study	10
Method	12
Participants	12
Measures	12
Development Center Design and Scoring Protocol	13
Procedure	14
Results	16
Discussion	27
Limitations and Implications	28
References	30
Appendix	34

Background

Millennia ago, the ancient Greek maxim “know thyself” was inscribed on the temple at Delphi. The wisdom of this aphorism advises that people “know” themselves—recognize their strengths and weaknesses and understand how others perceive their behavior. This type of self-awareness is especially germane for individuals employed in 21st century workplaces. In order to navigate the political and cultural realities of modern organizations, individuals must possess an accurate self-perception (Ostroff, Atwater, & Feinberg, 2004). One way to gauge the accuracy of a person’s self-perception is to compare that individual’s self-assessment against other people’s assessments of him or her. This technique, known as self-other rating agreement (SORA), is often included in organizations’ performance management practices, such as 360° feedback. It has been postulated that, over an individual’s career, declining SORA discrepancy could serve as an indicator of improved self-perception accuracy (Yammarino & Atwater, 1997).

SORA has utility outside of organizational environments as well. As first posited by Heneman (1980), the simulated and objective context of assessment centers offers an ideal setting in which to incorporate and study self-assessment and SORA. While some studies have embraced this suggestion and investigated self-assessment in assessment centers (e.g., Randall, Ferguson, & Patterson, 2000; Schmitt, Ford, & Stults, 1986), fewer have researched self-assessment in development centers (Halman & Fletcher, 2000; Jackson, Stillman, Burke, & Englert, 2007; Rupp, Baldwin, & Bashshur, 2006). To advance the literature on the latter, the present study investigates the effects and tendencies of participants’ self-assessments of their performance in a development center. Specifically, this research examines SORA of the various competencies and exercises assessed in the development center, considers gender differences, and compares participants’ performance based on their SORA categorization.

Development Centers

Derived from the well-known assessment center methodology popularized by Howard & Bray (1988), development centers have gained popularity in recent years (Kudisch et al., 2001). Rupp, Snyder, Gibbons, & Thornton (2006) define a development center as a “collection of workplace simulation exercises and other assessments that provide individuals with practice, feedback, and coaching on a set of developable behavioral dimensions found to be critical for their professional success” (p.78). While traditional assessment centers are diagnostic in nature and used to make personnel decisions, development centers emphasize experiential learning, participant self-reflection, feedback, and performance improvement on specific behavioral competencies (Rupp et al., 2006). Development centers are generally considered fair (e.g., exercises clearly relate to job tasks) and objective (e.g., assessment of actual behavior rather than psychological constructs) (Byham, 2002). Meta-analyses suggest that centers have criterion-related validity ranging from .28 to .37 (Gaugler, Rosenthal, Thornton, & Bentson, 1987; Hermelin, Lievens, & Robertson, 2007).

Self-Other Rating Agreement

Research and interest in SORA has grown since the 1997 publication of Atwater and Yammarino’s seminal paper on the topic. SORA is typically defined as the degree of agreement or congruence between one’s self-ratings and the ratings of others (Fleenor, Smither, Atwater, Braddy, & Sturm, 2010). Coworkers such as superiors, peers, and subordinates, and sometimes even customers or clients, provide the other ratings—typically through multisource (360°) feedback tools—for the purposes of performance appraisal or development (Yammarino & Atwater, 1997). Much of the current SORA research focuses on how it relates to various outcomes of interest, such as leadership performance and derailment. Accurate SORA has

instrumental value and is associated with positive performance, leadership, and organizational outcomes (Yammarino & Atwater, 1993), whereas individuals with inaccurate SORA may misdiagnose their managerial abilities and adversely impact their leadership (Fleenor, McCauley, & Brutus, 1996). Years before, Thornton (1980) concluded that individuals tend to view their own job performance very differently than other people do. So, while congruence between one's self-ratings and others' ratings is certainly encouraging and desirable, discrepancy is also salient for individuals because it illustrates how they are perceived by others.

SORA discrepancy can be largely attributed to what researchers have known for over a century: self-ratings are problematic. They are liable to be inflated, unreliable, biased, inaccurate (Podsakoff & Organ, 1986), and otherwise suspect when compared to other's ratings or objective measures. Self-ratings also tend to be higher than others' ratings (Atwater & Yammarino, 1992).

Factors Affecting SORA

Several factors, especially gender, cognitive ability, and context, have been shown to influence self-ratings and SORA. Men tend to rate themselves inaccurately higher compared to women, who generally have more congruent SORA (Yammarino & Atwater, 1997). In other words, women tend to have more accurate self-perceptions than men. Likewise, Brutus, Fleenor, and McCauley (1999) found that men tended to overestimate their managerial effectiveness, while women's self-ratings of it aligned with the ratings of others. In a selection context, men showed a tendency to inflate their self-ratings compared to women (Jones & Fletcher, 2002). However, an earlier study by Atwater and Roush (1994) found that self-ratings made confidentially showed no difference between genders. Most recently, Taylor, Sturm, Atwater, and Braddy (2016) offer evidence indicating that men and women will not differ in terms of SORA congruence. Such inconsistent findings warrant further research.

Yammarino and Atwater (1997) suggest higher intelligence, greater cognitive complexity, and better memories help individuals collect, process, and retain information, factors that consistently result in less inflated/more congruent SORA. Individuals with these cognitive traits possess the tools necessary to recognize and accurately report self-relevant information, such as their own thought processes and perceptions. Metacognitive ability also impacts SORA, as evidenced by Kruger and Dunning's (1999) fascinating finding that self-ratings of ability or performance were most grossly inaccurate for those individuals who scored the lowest on said ability or performance measure. In other words, these individuals not only performed poorly, but they also failed to recognize it. Unconscious biases can also impact SORA. Self-ratings may suffer from social desirability and leniency bias, as well as self-serving attribution bias—the tendency for self-raters to attribute their positive performance to stable personal factors and their negative performance to uncontrollable or contextual factors (Yammarino & Atwater, 1993).

A range of contextual factors affect SORA. Among them are the environment or situation in which self-ratings are obtained (e.g., development centers, assessment centers, or performance reviews) and political or personnel purposes, such as whether self-ratings are being obtained for selection, evaluation, or development decisions (Fleenor et al., 2010). For example, Heidemeier and Moser (2009) found that managers tended to overrate their own ability in a developmental performance context—an unexpected effect perhaps explained by ambivalence or a lack of pressure felt by the managers to self-rate accurately in the unrewarding developmental setting.

SORA in Centers

SORA has previously been studied, to varying degrees, in assessment center and development center settings. Most studies focused on SORA not as an individual construct but as a predictor or criterion variable related to some other construct or research question. For

example, Randall et al. (2000) tested the relationship between the ability to produce accurate SORA and the outcome of an assessment center. They found that accurate SORA consistently discriminated between accepted and rejected candidates (e.g., rejected candidates consistently overestimated their performance on exercise and had larger SORA discrepancies than accepted candidates). Atkins and Wood (2002) corroborated Randall et al.'s (2000) findings. Comparing self-ratings from a multi-rater instrument to assessment center performance, Atkins and Wood (2002) found that participants who rated themselves the highest on the instrument were the worst performers in the assessment center. They concluded that self-ratings were negatively and nonlinearly related to performance.

Jackson, Stillman, Burke, and Englert (2007) considered SORA in conjunction with social dominance and cognitive ability measures in a development center. Using cluster analysis, they found that dominance-related personality dimensions differentiated among individuals with congruent and incongruent SORA. Namely, individuals who inflated their self-ratings also had higher social dominance orientations.

In another usage of SORA, Schmitt, Ford, and Stults (1986) had assessment center participants self-rate on eight competencies both before and after the center. These self-ratings were later correlated with assessor ratings. Results showed significant changes in self-ratings on five of the eight competencies. The authors also reported exercises and competencies on which self-assessment changes were observed. Halman and Fletcher (2000) used the same pre-post center methodology to investigate self-assessment, this time in a development center. Before the center, SORA congruence was found on only two of the ten competencies being assessed. After the center, SORA congruence was found to have risen to six of the ten competencies. The

authors also reported that women had more accurate SORA than men and that development centers may be useful for increasing self-awareness, operationalized as SORA.

In another development center study, Rupp, Baldwin, and Bashshur (2006) found congruent SORA almost across the board. Observers' assessments of participants' performance generally concurred with the participants' self-ratings of performance on competencies such as information seeking, oral communication, planning and organizing, problem solving, and conflict management.

Present Study

The present study intends to: 1) adopt Heneman's (1980) notion that assessment centers are optimal settings in which to study self-assessment, 2) advance the research on self-assessment across assessment methods and contexts, including development, as called for by Randall et al. (2000), 3) remedy Halman and Fletcher's (2000) inability to examine how exercises impact self-assessments, 4) focus on self-assessment and gender, as Halman and Fletcher (2000) did, and 5) compare results to other findings about which competencies and exercises appear to affect self-assessment (Schmitt, Ford, & Stults, 1986; Jackson, Stillman, Burke, & Englert, 2007). In short, the purpose of this present study is to investigate the effects and tendencies of participants' self-assessments of performance in a development center. The following hypotheses and questions are founded on previous research on self-assessment accuracy, development center performance, and self-other rating agreement.

Hypothesis 1. Participants' self-ratings will be higher than other-ratings.

Hypothesis 2. The difference between self-ratings and other-ratings will be smaller for women than men.

Question 3a. Which exercises result in congruent SORA?

Hypothesis 3b. There will be gender differences on which exercises result in congruent SORA.

Question 4a. Which competencies result in congruent SORA?

Hypothesis 4b. There will be gender differences on which competencies result in congruent SORA.

Hypothesis 5. Participants who completed self-ratings will receive higher other-ratings than participants who did not complete self-ratings.

Hypothesis 6. Participants who underrated themselves will receive higher other-ratings than participants who overrated or accurately self-rated.

Hypothesis 7. Participants who overrated themselves will receive lower other-ratings than participants who underrated or accurately self-rated.

Method

Participants

Participants were 139 executives (70 men, 69 women) from 13 development centers conducted between June 2011 and August 2016. At the time of their involvement, all participants were junior level executives employed by a large global biotechnology company. For 6 centers, the company opted for a design that did not include self-assessment for the participants ($n=61$). The other 7 centers included a self-assessment component for participants ($n=78$). See Table 1.

Table 1

<i>Development Center Participants</i>			
Center	Men	Women	Total
Jun 2011*	6	6	12
Aug 2011*	5	6	11
Nov 2011*	6	5	11
Feb 2012*	4	7	11
Aug 2012*	5	4	9
Oct 2012	6	2	8
Apr 2013	7	5	12
Jun 2013	3	8	11
Aug 2013	7	5	12
Oct 2013	3	5	8
Mar 2014	6	4	10
May 2016*	6	6	12
Aug 2016*	6	6	12
Total	70	69	139

* denotes centers including self-assessment

Measures

Participants were assessed on several exercises and leadership competencies. Each center included most of the same exercises and competencies. Exercises and competencies on which a

minority ($n < 104$) of participants were assessed were excluded from analyses. The exercises and competencies on which a majority ($n > 103$) of participants were assessed are listed in Table 2. Rating forms for the six exercises can be found in the appendix. Content validity is the primary source of validity available for the measures used in the development centers. However, concurrent criterion-related validity studies conducted on development centers with a similar design found them to be significant predictors of performance with effect sizes of .48 to .60.

Table 2

<i>Development Center Exercises and Competencies</i>	
Exercises	Competencies
Strategy Discussion I	Communication
Employee Conversation	Decision Making
Strategy Discussion II	Teamwork & Collaboration
Town Hall Presentation	Achieving Results
Conflict Resolution	Innovation
Board Presentation	Managing Change
	Inspiring & Influencing

Development Center Design and Scoring Protocol

The centers were designed to measure performance at a level above (i.e., senior executive) that of the participants' current position. Centers were conducted over a four-day period with three days of assessment and one day dedicated to feedback and development planning. Exercises consisted of leaderless group discussions, presentations (based on a business case), roleplays, and a behaviorally based interview. Three to four leadership competencies were assessed in each exercise.

There were at least three observers per exercise. Observer teams consisted of one internal and two external individuals. All observers participated in a half-day training session prior to

each center. Observers rotated such that a participant's ratings were derived from multiple assessments by different observers in multiple exercises. During each exercise, observers consulted behavioral marker guides and indicated if they observed the participant exhibit a demarcated behavior during said exercise. After each exercise, observers independently rated the participant's proficiency on specific competencies on a scale of 1 (learning/development area) to 5 (exceptional). Observers then convened, discussed, and adjusted their independent ratings to establish consistency (i.e., all ratings within one point of each other). Interrater reliability statistics are unavailable as a result of this consensus rating process. Observer ratings were captured for every competency measured within each exercise; however, no exercise assessed every competency. Observer ratings were aggregated and averaged to determine an overall rating for each exercise and an overall rating for each competency.

After completing each exercise, participants reflected on and rated their own performance on the same competencies using the same scale as the observers. Thus, participant self-ratings were captured for each competency within each exercise. An overall self-rating for each exercise and each competency was also calculated for every participant. However, as mentioned, 6 centers did not include participant self-ratings.

Procedure

Data from each of the 13 centers were extracted from 13 separate transactional databases and cleaned, merged, and sorted into one master database. This database contains observer ratings and participant self-ratings (when available) for each exercise and competency. The database also includes the average rating of all exercises combined from both observers and participants. The intent of this present study is to use the observer ratings from each exercise and

competency to establish a performance baseline against which to compare participants' self-ratings and examine the degree of agreement or congruence between the two sets of ratings.

Results

Hypothesis 1. Participants' self-ratings will be higher than other-ratings.

Hypothesis 1 proposed that participants' self-ratings would be higher than their corresponding observer ratings. Results of paired samples t-tests provided minimal support for this hypothesis. Analyzed collectively, participants' self-ratings were significantly higher than their corresponding observer ratings for only 2 of 6 exercises (Strategy Discussion I and Board Presentation) and 2 of 7 competencies (Teamwork & Collaboration and Innovation). See Tables 3 and 4.

Table 3

Participant Against Observer Ratings of Exercises

	Participants			Observers			<i>df</i>	<i>t</i>
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>		
Strategy Discussion I	78	2.31	.595	78	2.10	.433	77	2.75**
Employee Conversation	77	2.26	.719	77	2.12	.607	76	1.49
Strategy Discussion II	78	2.20	.562	78	2.10	.416	77	1.71
Town Hall Presentation	77	1.93	.611	77	2.08	.619	76	-1.72
Conflict Resolution	54	2.10	.621	54	2.19	.694	53	-.916
Board Presentation	44	2.22	.595	44	1.97	.433	43	2.78**

* $p \leq 0.05$ ** $p \leq 0.01$

Table 4

Participant Against Observer Ratings of Competencies

	Participants			Observers			<i>df</i>	<i>t</i>
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>		
Communication	78	2.34	.551	78	2.31	.412	77	.427
Decision Making	78	2.17	.622	78	2.10	.486	77	.864
Teamwork & Collaboration	78	2.51	.655	78	2.25	.459	77	3.33**
Achieving Results	78	2.01	.559	78	2.11	.500	77	-1.27
Innovation	78	2.07	.638	78	1.78	.398	77	4.17**

Managing Change	78	1.95	.558	78	1.92	.508	77	.379
Inspiring & Influencing	78	2.20	.559	78	2.08	.459	77	1.58

* $p \leq 0.05$ ** $p \leq 0.01$

Gender breakdowns revealed starker differences between self-ratings and observer ratings. Results of paired samples t-tests showed that men had significantly higher self-ratings than observer ratings for 4 of 6 exercises (Strategy Discussion I, Employee Conversation, Strategy Discussion II, and Board Presentation) and 4 of 7 competencies (Teamwork & Collaboration, Innovation, Managing Change, and Inspiring & Influencing). See Tables 5 and 6. Men also had higher self-ratings than observer ratings for the 2 other exercises (Town Hall Presentation and Conflict Resolution) and the 3 other competencies (Communication, Decision Making, and Achieving Results), but the group differences were not statistically significant.

Table 5

Male Participant Against Observer Ratings of Exercises

	Men			Observers			<i>df</i>	<i>t</i>
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>		
Strategy Discussion I	38	2.51	.574	38	2.17	.404	37	3.04**
Employee Conversation	37	2.48	.725	37	2.11	.536	36	2.74**
Strategy Discussion II	38	2.36	.600	38	2.15	.372	37	2.46*
Town Hall Presentation	38	2.07	.575	38	2.00	.500	37	.758
Conflict Resolution	26	2.32	.517	26	2.12	.643	25	1.44
Board Presentation	21	2.38	.562	21	2.09	.431	20	2.34*

* $p \leq 0.05$ ** $p \leq 0.01$

Table 6

Male Participant Against Observer Ratings of Competencies

	Men			Observers			<i>df</i>	<i>t</i>
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>		
Communication	38	2.47	.559	38	2.30	.347	37	1.94
Decision Making	38	2.32	.563	38	2.19	.448	37	1.20
Teamwork & Collaboration	38	2.70	.693	38	2.30	.421	37	3.43**
Achieving Results	38	2.20	.536	38	2.11	.450	37	.822
Innovation	38	2.33	.690	38	1.84	.363	37	4.75**
Managing Change	38	2.11	.552	38	1.87	.442	37	2.44*
Inspiring & Influencing	38	2.39	.506	38	2.03	.358	37	3.98**

* $p \leq 0.05$ ** $p \leq 0.01$

Results of paired samples t-tests indicated that women, as opposed to men, self-rated significantly lower than their corresponding observer ratings on 2 of 6 exercises (Town Hall Presentation and Conflict Resolution) and 1 of 7 competencies (Achieving Results). See Tables 7 and 8. However, women self-rating lower than observers was not necessarily the case for the other exercises and competencies.

Table 7

Female Participant Against Observer Ratings of Exercises

	Women			Observers			<i>df</i>	<i>t</i>
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>		
Strategy Discussion I	40	2.13	.563	40	2.04	.456	39	.873
Employee Conversation	40	2.06	.659	40	2.13	.672	39	-.600
Strategy Discussion II	40	2.04	.479	40	2.04	.452	39	-.058
Town Hall Presentation	39	1.79	.620	39	2.17	.712	38	-2.70**
Conflict Resolution	28	1.89	.647	28	2.25	.745	27	-2.60*
Board Presentation	23	2.07	.595	23	1.86	.415	22	1.60

* $p \leq 0.05$ ** $p \leq 0.01$

Table 8

Female Participant Against Observer Ratings of Competencies

	Women			Observers			<i>df</i>	<i>t</i>
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>		
Communication	40	2.21	.519	40	2.32	.471	39	-1.24
Decision Making	40	2.03	.650	40	2.01	.510	39	.116
Teamwork & Collaboration	40	2.34	.571	40	2.20	.493	39	1.27
Achieving Results	40	1.84	.530	40	2.10	.548	39	-2.46*
Innovation	40	1.83	.474	40	1.73	.426	39	1.17
Managing Change	40	1.79	.525	40	1.96	.565	39	-1.58
Inspiring & Influencing	40	2.01	.550	40	2.13	.538	39	-1.21

* $p \leq 0.05$ ** $p \leq 0.01$

Hypothesis 2. The difference between self-ratings and other-ratings will be smaller for women than men.

Hypothesis 2 posited that women would self-rate more accurately than men. Results of analyses generally support this hypothesis. Refer to the results of paired samples t-tests presented in Tables 5, 6, 7, and 8. Men did not self-rate significantly differently (although their ratings were indeed numerically higher) from their corresponding observer ratings for only 2 of 6 exercises (Town Hall Presentation and Conflict Resolution) and only 3 of 7 competencies (Communication, Decision Making, and Achieving Results). These results suggest that, for men, inaccurate SORA occurred more often than not. See Tables 5 and 6.

Women, on the other hand, did not self-rate significantly differently from their corresponding observer ratings for 4 of 6 exercises (Strategy Discussion I, Employee Conversation, Strategy Discussion II, and Board Presentation) and 6 of 7 competencies (all but Achieving Results). These results suggest that, for women, accurate SORA occurred more often

than not. See Tables 7 and 8. Again, it was not necessarily the case that women self-rated lower than observers for the other exercises and competencies.

Additionally, results of two-tailed bivariate Pearson's correlations indicated that men had SORA on 2 of 6 exercises (Strategy Discussion II and Town Hall Presentation) and 2 of 7 competencies (Communication and Innovation), while women had SORA on 3 of 6 exercises (Employee Conversation, Strategy Discussion II, and Conflict Resolution) and 3 of 7 competencies (Communication, Innovation, and Inspiring & Influencing). See Tables 10 and 14.

Question 3a: Which exercises result in congruent SORA?

Research question 3a sought to determine which exercises resulted in the most congruent SORA. Results of two-tailed bivariate Pearson's correlations indicated that, collectively, participants had SORA on 4 of 6 exercises (Employee Conversation, Strategy Discussion II, Conflict Resolution, and Board Presentation). See Table 9.

Table 9

Self-Other Rating Agreement of Exercises

	<i>n</i>	<i>r</i>
Strategy Discussion I	78	.15
Employee Conversation	77	.24*
Strategy Discussion II	78	.49**
Town Hall Presentation	77	.21
Conflict Resolution	54	.34*
Board Presentation	44	.37*

* $p \leq 0.05$ ** $p \leq 0.01$

Hypothesis 3b: There will be gender differences on which exercises result in congruent SORA.

Hypothesis 3b posited that gender differences would exist in terms of SORA among exercises. Results of the analyses did little to support this hypothesis. Two-tailed bivariate Pearson's correlations indicated that men had SORA on 2 of 6 exercises (Strategy Discussion II

and Town Hall Presentation), and women had SORA on 3 of 6 exercises (Employee Conversation, Strategy Discussion II, and Conflict Resolution). See Table 10.

Results of two-tailed independent samples t-tests revealed that, except for the Board Presentation exercise, men self-rated significantly higher than women on all exercises. See Table 11. There were no significant differences between men and women in terms of their corresponding observer ratings on all exercises, however. See Table 12.

Table 10

Self-Other Rating Agreement of Exercises by Gender

	Men		Women	
	<i>n</i>	<i>r</i>	<i>n</i>	<i>r</i>
Strategy Discussion I	38	.05	40	.17
Employee Conversation	37	.17	40	.32*
Strategy Discussion II	38	.51**	40	.45**
Town Hall Presentation	38	.38*	39	.17
Conflict Resolution	26	.33	28	.45*
Board Presentation	21	.36	23	.30

* $p \leq 0.05$ ** $p \leq 0.01$

Table 11

Participant Self-Ratings of Exercises by Gender

	Men			Women			<i>df</i>	<i>t</i>
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>		
Strategy Discussion I	38	2.51	.574	40	2.13	.563	76	2.92**
Employee Conversation	37	2.48	.725	40	2.06	.659	75	2.69**
Strategy Discussion II	38	2.36	.600	40	2.04	.479	76	2.64**
Town Hall Presentation	38	2.07	.575	39	1.79	.620	75	2.04*
Conflict Resolution	26	2.32	.517	28	1.89	.647	52	2.65*
Board Presentation	21	2.38	.562	23	2.07	.595	42	1.81

* $p \leq 0.05$ ** $p \leq 0.01$

Table 12

Observer Ratings of Exercises by Participant Gender

	Men			Women			<i>df</i>	<i>t</i>
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>		
Strategy Discussion I	70	2.21	.400	69	2.10	.427	137	1.66
Employee Conversation	69	2.08	.535	68	2.10	.668	135	-.195
Strategy Discussion II	70	2.19	.375	68	2.07	.450	136	1.66
Town Hall Presentation	70	2.09	.580	67	2.17	.664	135	-.837
Conflict Resolution	58	2.27	.650	56	2.15	.714	112	.927
Board Presentation	53	2.10	.464	51	2.01	.499	102	.979

* $p \leq 0.05$ ** $p \leq 0.01$

Question 4a: Which competencies result in congruent SORA?

Research question 4a sought to determine which competencies resulted in the most congruent SORA. Results of two-tailed bivariate Pearson's correlations indicated that, collectively, participants had SORA on 5 of 7 competencies (Communication, Decision Making, Teamwork & Collaboration, Innovation, and Inspiring & Influencing). See Table 13.

Table 13

Self-Other Rating Agreement of Competencies

	<i>n</i>	<i>r</i>
Communication	78	.33**
Decision Making	78	.23*
Teamwork & Collaboration	78	.24*
Achieving Results	78	.21
Innovation	78	.39**
Managing Change	78	.18
Inspiring & Influencing	78	.23*

* $p \leq 0.05$ ** $p \leq 0.01$

Hypothesis 4b: There will be gender differences on which competencies result in congruent SORA.

Hypothesis 4b posited that gender differences would exist in terms of SORA among competencies. Results of the analyses did little to support this hypothesis. Two-tailed bivariate Pearson's correlations indicated that men had SORA on 2 of 7 competencies (Communication and Innovation), and women had SORA on 3 of 7 competencies (Communication, Innovation, and Inspiring & Influencing). See Table 14.

Results of two-tailed independent samples t-tests revealed that men self-rated significantly higher than women on all competencies. See Table 15. In terms of their corresponding observer ratings, men were rated higher than women on the Decision Making competency, but there were no significant differences between men and women on any other competency. See Table 16.

Table 14

Self-Other Rating Agreement of Competencies by Gender

	Men		Women	
	<i>n</i>	<i>r</i>	<i>n</i>	<i>r</i>
Communication	38	.34*	40	.36*
Decision Making	38	.20	40	.19
Teamwork & Collaboration	38	.23	40	.22
Achieving Results	38	.24	40	.20
Innovation	38	.42**	40	.34*
Managing Change	38	.25	40	.19
Inspiring & Influencing	38	.21	40	.34*

* $p \leq 0.05$ ** $p \leq 0.01$

Table 15

Participant Self-Ratings of Competencies by Gender

	Men			Women			<i>df</i>	<i>t</i>
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>		
Communication	38	2.47	.559	40	2.21	.519	76	2.18*
Decision Making	38	2.32	.563	40	2.03	.650	76	2.11*
Teamwork & Collaboration	38	2.70	.693	40	2.34	.571	76	2.51*
Achieving Results	38	2.20	.536	40	1.84	.530	76	2.95**
Innovation	38	2.33	.690	40	1.83	.474	76	3.77**
Managing Change	38	2.11	.552	40	1.79	.525	76	2.61*
Inspiring & Influencing	38	2.39	.506	40	2.01	.550	76	3.15**

* $p \leq 0.05$ ** $p \leq 0.01$

Table 16

Observer Ratings of Competencies by Participant Gender

	Men			Women			<i>df</i>	<i>t</i>
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>		
Communication	70	2.31	.349	68	2.32	.484	136	-.141
Decision Making	70	2.25	.443	69	2.06	.454	137	2.42*
Teamwork & Collaboration	70	2.30	.428	69	2.26	.468	137	.542
Achieving Results	70	2.16	.401	68	2.07	.504	136	1.09
Innovation	70	1.92	.355	69	1.80	.440	137	1.78
Managing Change	70	1.97	.454	68	1.98	.521	136	-.141
Inspiring & Influencing	70	2.10	.403	68	2.10	.545	136	-.072

* $p \leq 0.05$ ** $p \leq 0.01$

Hypothesis 5: Participants who completed self-ratings will receive higher other-ratings than participants who did not complete self-ratings.

Hypothesis 5 proposed that participants who completed self-ratings would receive higher observer ratings than participants who did not complete self-ratings. Results of several two-tailed independent samples t-tests did not support this hypothesis. In terms of observer ratings on

exercises, there were no significant differences between participants who self-rated and participants who did not self-rate. Only on the Innovation competency were participants who did not self-rate rated higher than participants who did self-rate. There were no significant differences between the two groups on any other competencies. See Tables 17 and 18.

Table 17

Observer Ratings of Exercises by Participant Self-Evaluation

	Self-Evaluation			No Self-Evaluation			<i>df</i>	<i>t</i>
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>		
Strategy Discussion I	78	2.10	.433	61	2.22	.386	137	1.72
Employee Conversation	77	2.12	.607	60	2.05	.600	135	.673
Strategy Discussion II	78	2.10	.416	60	2.17	.416	136	1.00
Town Hall Presentation	77	2.08	.619	60	2.19	.627	135	-.966
Conflict Resolution	54	2.19	.694	60	2.23	.676	112	-.329
Board Presentation	44	1.97	.433	60	2.12	.507	102	1.65

* $p \leq 0.05$ ** $p \leq 0.01$

Table 18

Observer Ratings of Competencies by Participant Self-Evaluation

	Self-Evaluation			No Self-Evaluation			<i>df</i>	<i>t</i>
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>		
Communication	78	2.31	.412	60	2.33	.431	136	-.351
Decision Making	78	2.10	.486	61	2.23	.408	137	1.69
Teamwork & Collaboration	78	2.25	.459	61	2.31	.433	137	-.826
Achieving Results	78	2.11	.500	60	2.13	.393	136	-.258
Innovation	78	1.78	.398	61	1.96	.390	137	2.61**
Managing Change	78	1.92	.508	60	2.05	.450	136	1.55
Inspiring & Influencing	78	2.08	.459	60	2.12	.503	136	-.432

* $p \leq 0.05$ ** $p \leq 0.01$

Note: It was determined that the current study's sample size was insufficient to conduct analyses to investigate **Hypothesis 6** (Participants who underrated themselves will receive higher

other-ratings than participants who overrated or accurately self-rated) and *Hypothesis 7* (Participants who overrated themselves will receive lower other-ratings than participants who underrated or accurately self-rated). See discussion of this limitation further below.

Discussion

The most notable findings from this study relate to differences between men and women regarding self-ratings and others' ratings of their development center performance. Comparing participants' self-ratings to their corresponding observers' ratings helps indicate a degree of congruence or incongruence between the two sets of ratings. The present study found that men rated themselves significantly higher than observers for 4 of 6 exercises and 4 of 7 competencies. Women, on the other hand, did not rate themselves significantly higher than observers on any exercises or competencies. In fact, women rated themselves significantly lower than observers on 2 of 6 exercises and 1 of 7 competencies. The stark gender differences between self-ratings and observer ratings lend credence to findings from Halman and Fletcher (2000) and Rupp, Baldwin, and Bashshur (2006) on gender and SORA. For whatever reasons (too numerous and complex for the scope of this paper), it seems women undervalue their performance and men overestimate theirs. These rating tendencies may be attributable to gender differences on constructs like humility, self-confidence, and self-awareness, but this suggestion is speculative.

Men and women also differed from each other in terms how they self-rated their performance on exercises and competencies. Except for the Board Presentation exercise, men self-rated significantly higher than women on all exercises. Likewise, men self-rated significantly higher than women on all competencies. These results corroborate findings from Yammarino and Atwater (1997), Brutus, Fleenor, and McCauley (1999), and Jones and Fletcher (2002) that suggested men tend to rate themselves higher compared to women (i.e., compared to women, men inflated their self-ratings). Although findings related to gender differences, self-ratings, and SORA remain inconsistent, results from the present study add to the body of literature supporting the idea that gender differences exist within those constructs.

Another dichotomy of interest in this study compared participants who completed self-evaluations to participants who did not. The finding that self-evaluation appeared to have no effect on observer ratings one way or the other calls into question the utility of self-evaluation practices as part of the development center model. One might hope that self-evaluation practices would heighten participants' metacognition and enhance their subsequent behaviors in the development center, but the results of this study suggest that is not the case. Despite this finding, self-evaluation may still have practical utility within the development center model. If nothing else, it serves as a rich source of personal information to be discussed in executive coaching sessions or performance management meetings. Further research is needed to investigate the value of self-evaluation in development centers. A time series design may be an appropriate analysis for such future research.

Limitations and Implications

As previously mentioned, this study's sample size precluded analyses to investigate Hypothesis 6 and Hypothesis 7. To do so would have required self-ratings and observer ratings from substantially more participants than the ≤ 78 included in this study. Furthermore, the three rater categories proposed in the hypotheses (underrater, overrater, and accurate rater) would not capture all the distinct rater categories that inevitably emerge during such analyses. Using anywhere from 5 to 12 rater categories, such as extreme underrater, extreme overrater, accurate underrater, and accurate overrater—as well as the three aforementioned categories—involve breakdowns that necessitate a larger sample size.

Categorization issues aside, the sample size of this study was still undesirably small. While data (of various amounts) were available for 139 participants, most analyses included 40

or fewer participants, due to the breakdowns by gender, exercise, competency, or self-evaluation needed to test the stated hypotheses.

Obviously, the limitations described here could be largely remedied by an increased sample size, provided data for those participants came from development centers incorporating identical exercises, competencies, and self-evaluation practices.

While results of this study do little to advance the theory behind SORA, they still contribute to existing literature about gender, self-assessment, and development center exercises and competencies. If nothing else, perhaps merely knowing about the effects and tendencies of development center performance as described in this study could help observers, coaches, and participants enhance their self-awareness, thus rendering them better able to “know thyself.”

References

- Atkins, P.W.B., & Wood, R.E. (2002). Self- versus others' ratings as predictors of assessment center ratings: Validation evidence for 360-degree feedback programs. *Personnel Psychology, 55*, 871-903.
- Atwater, L.E., & Roush, P.E. (1994). An investigation of gender effects on followers' ratings of leaders, leaders' self-ratings and reactions to feedback. *The Journal of Leadership Studies, 1*(4), 37-52.
- 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. In G.R. Ferris (Ed.), *Research in personnel and human resources management, 15*. (pp. 121-174). JAI Press.
- Brutus, S., Fleenor, J.W., & McCauley, C.D. (1999). Demographic and personality predictors of congruence in multi-source ratings. *The Journal of Management Development, 18*, 417-435.
- Byham, W.C. (2002). What is an assessment center? The assessment center method, applications, and technologies. Development Dimensions International, Inc.
- Fleenor, J.W., Smither, J.W., Atwater, L.E., Braddy, P.W., & Sturm, R.E. (2010). Self-other rating agreement in leadership: A review. *Leadership Quarterly, 21*, 1005-1034.
- Fleenor, J.W., McCauley, C.D., & Brutus, S. (1996). Self-other rating agreement and leader effectiveness. *Leadership Quarterly, 7*(4), 487-506.
- Gaugler, B.B., Rosenthal, D.B., Thornton, G.C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology, 72*, 493-511.

- Halman, F., & Fletcher, C. (2000). The impact of development centre participation and the role of individual differences in changing self-assessments. *Journal of Occupational and Organizational Psychology*, 73, 423-442.
- Heidemeier, H., & Moser, K. (2009). Self-other agreement in job performance ratings: A meta-analytic test of a process model. *Journal of Applied Psychology*, 94, 353-370.
- Heneman, H.G. (1980). Self-assessment: A critical analysis. *Personnel Psychology*, 33(2), 297-300.
- Hermelin, E., Lievens, F., & Robertson, I.T. (2007). The validity of assessment centres for the prediction of supervisory performance ratings: A meta-analysis. *International Journal of Selection and Assessment*, 15, 405-411.
- Howard, A., & Bray, D.W. (1988). *Managerial lives in transition: Advancing age and changing times*. New York: Guilford.
- Jackson, D.J.R., Stillman, J.A., Burke, S., & Englert, P. (2007). Self versus assessor ratings and their classification in assessment centres: Profiling the self-rater. *New Zealand Journal of Psychology*, 36(2), 93-99.
- Jones, L., & Fletcher, C. (2002). Self-assessment in a selection situation: An evaluation of different measurement approaches. *Journal of Occupational & Organizational Psychology*, 75, 145-161.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121-1134.
- Kudisch, J.D., Avis, J.M., Fallon, J.D., Thibodeaux, H.F., Roberts, F.E., Rollier, T.J., et al. (2001, April). *A survey of assessment center practices in organizations worldwide:*

- Maximizing innovation or business as usual?* Paper presented at the 16th annual conference for the Society of Industrial and Organizational Psychology, San Diego, CA.
- Ostroff, C., Atwater, L.E., & Feinberg, B.J. (2004). Understanding self-other agreement: A look at rater and rate characteristics, context, and outcomes. *Personnel Psychology, 57*, 333-375.
- Podsakoff, P.M., & Organ, D.W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management, 12*(4), 531-544.
- Randall, R., Ferguson, E., & Patterson, F. (2000). Self-assessment accuracy and assessment centre decisions. *Journal of Occupational and Organizational Psychology, 73*(4), 443-459.
- Rupp, D.E., Baldwin, A., & Bashshur, M. (2006). Using developmental assessment centers to foster workplace fairness. *The Psychologist-Manager Journal, 9*(2), 145-170.
- Rupp, D.E., Snyder, L.A., Gibbons, A.M., & Thornton III, G.C. (2006). What should developmental assessment centers be developing? *The Psychologist-Manager Journal, 9*(2), 75-98.
- Schmitt, N., Ford, J.K., & Stults, D.M. (1986). Changes in self-perceived ability as a function of performance in an assessment centre. *Journal of Occupational Psychology, 59*, 327-335.
- Taylor, S.N., Sturm, R.E., Atwater, L.E., & Braddy, P.W. (2016). Underestimating one's leadership impact: Are women leaders more susceptible? *Organizational Dynamics, 45*, 132-138.
- Thornton, G.C. (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology, 33*, 263-271.

Yammarino, F.J., & Atwater, L.E. (1997). Do managers see themselves as others see them?

Implications of self-other rating agreement for human resources management.

Organizational Dynamics, 35-44.

Yammarino, F.J., & Atwater, L.E. (1993). Understanding self-perception accuracy: Implications

for human resource management. *Human Resource Management*, 32, 231-247.

Appendix

Strategy Discussion I – Feedback Form					Participant: Observer:	
Strategy Setting/Professional & Business Expertise (Average Rating)						
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>		
Strength			Area of Improvement			
Decision Making (Average Rating)						
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>		
Strength			Area of Improvement			
Teamwork & Collaboration (Average Rating)						
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>		
Strength			Area of Improvement			
Innovation (Average Rating)						
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>		
Strength			Area of Improvement			

Employee Conversation – Feedback Form					Participant: Observer:
Inspiring & Influencing (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		
Communication (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		
Enabling Performance (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		
Coaching & Development (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		

Strategy Discussion II – Feedback Form					Participant: Observer:
Strategy Setting/ Professional & Business Expertise (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		
Decision Making (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		
Teamwork & Collaboration (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		
Innovation (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		

Town Hall Presentation – Feedback Form					Participant: Observer:
Inspiring & Influencing (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		
Communication (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		
Achieving Results (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		
Managing Change (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		

Conflict Resolution – Feedback Form					Participant: Observer:
Inspiring & Influencing (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		
Communication (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		
Achieving Results (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		
Managing Change (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		

Board Presentation – Feedback Form					Participant: Observer:
Strategy Setting / Professional & Business Expertise (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		
Communication (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		
Achieving Results (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		
Managing Change (Average Rating)					
Learning/Dev. <input type="radio"/>	Basic <input type="radio"/>	Proficient <input type="radio"/>	Advanced <input type="radio"/>	Exceptional <input type="radio"/>	
Strength			Area of Improvement		