

A STUDY OF RUSSIAN, POLISH, AND U.S. SELF- AND PEER RATERS: MODERATORS OF RATING AGREEMENT

Robert D. Costigan
Richard C. Insinga
Grazyna Kranas
Selim S. Ilter
Vladimir A. Kureshov
J. Jason Berman

ABSTRACT

This study investigates one aspect of the multi-source feedback process: the agreement between self-ratings and coworker ratings of workplace behavior. Moderators of rating agreement (i.e., number of years that the coworker had known the employee, trustworthiness of the employee, and country status) are carefully examined. Eighty-six Russian employee-coworker dyads, 99 Polish dyads, and 95 U.S. dyads from more than 225 organizations participated. Regression results indicate that rating agreement was higher when the Polish and U.S. coworker knew the target employee a shorter period of time and when the Polish, Russian, and U.S. target employee was considered trustworthy.

INTRODUCTION

During the past decade, multi-source feedback has been introduced in many U.S. corporations (Mersman & Donaldson, 2000) and in firms located in the United Kingdom (Fletcher & Baldry, 2000). This employee-development strategy provides the employee with performance-related feedback from the self, supervisor, peers, subordinates, internal and external customers, and/or vendors. Ratings completed on one's own performance or behavior, otherwise known as self-ratings, are intended to more fully engage the target appraisee in the multi-source feedback process. Among the other raters involved in this process, coworkers may represent one of the more objective feedback sources for the target employee (Brutus, Fleenor, & McCauley, 1999). Coworkers normally have ample opportunity to observe their peers' workplace behavior without pretense, whereas the supervisor and subordinates may pose more of a threat to the target employee because of the hierarchical nature of their relationships. Thus, the ratings made by peers may be a valid predictor of an employee's actual performance. The present study closely investigates one aspect of self- and peer ratings, the extent to which these ratings agree. Possible moderators of rating agreement between these rating sources are also examined.

In the absence of "true" performance scores for the employee, which is typically the case

in field appraisal studies, rater accuracy has often been assessed with a proxy measure: the extent to which rating sources agree (Yammarino & Atwater, 1997). Even though agreement between different rating sources does not guarantee accuracy, accurate self- and peer ratings, in Yammarino and Atwater's view, are typically characterized by high interrater agreement.

Conway and Huffcutt's (1997) meta-analysis results indicate that the mean correlation between self-ratings and peer ratings is disappointingly low at .19, indicating that rating agreement historically has been low for these sources. By determining those rater characteristics and relationship characteristics that lead to increased agreement between self- and coworker ratings, it would be possible to improve this rating process. Therefore, the main purpose of this study is to identify moderators of rating agreement between self-raters and peer raters.

The raters in the present study come from two Eastern European countries in economic transition, Russia and Poland, and from the U.S., an established economy. Russia is of interest because of the business opportunities in its fast growing economy (Bush, 2003), and Poland is of interest because of its potential for economic growth in the European Union (Condon & Butler, 2003). Furthermore, the increasing presence of Western European and U.S. multinationals in Russia (Fey, Engstrom, & Bjorkman, 1999) and in Poland (Kozminski, 1999) makes it important to determine the applicability of Western human-resource practices, such as multi-source feedback, in subsidiaries located in these countries. HR scholars such as Kiriazov, Sullivan, & Tu (2000) have also encouraged the use of multiple rating sources in native Eastern European organizations. Assessing the agreement of Russian and Polish self- and peer raters seems to be worthwhile considering the potential for application of multi-source feedback systems in these countries. Possible moderators of rating agreement in these countries will be discussed next.

Time in a Relationship

In an earlier study of rating agreement, Brief, Aldag, and Van Sell (1977) found the correlation between the self-rating and the supervisor's rating of work quality to be higher when the self-rater had less job tenure. In a more recent study, Shore and Bleicken (1991) reported that middle-aged self-raters agreed less with their supervisor on two of six performance ratings when compared with the other two age groups in their study. Warr and Bourne (1999) found that rating congruence between the target employee and the supervisor was greater for younger subordinates than it was for older subordinates. Each of these studies looked at the moderating effects of time (i.e., tenure or age) on rating agreement between the self-rater and the supervisor. In our study, we plan to compare self-ratings with coworker ratings. In place of job tenure or age, another time-based variable (i.e., time in the relationship) has been incorporated into the design of this study. The duration of a peer relationship may be a more suitable time-based measure because it reflects the amount of time that the peer has had to observe the target employee's workplace behavior. In line with Brief et al. (1977) and Warr and Bourne's (1999) findings that the self-ratings made by less tenured or younger workers agree more with supervisory ratings, we put forth the following hypothesis for ratings involving coworkers:

Hypothesis 1: Agreement between self- and coworker ratings of workplace behavior will be significantly higher when these employees have known each other a shorter period of time.

One explanation proposed by Atwater and Yammarino (1997) for increased rating agreement between younger, less-tenured self-raters and their supervisors, which is the basis for

this study's first hypothesis involving self- and coworker ratings, is that less experienced workers are thought to be more open to receiving performance feedback than are more experienced workers. The comments coming from bosses may be taken more seriously by younger, less secure workers, continually causing them to reevaluate their own performance. Thus, the self-appraisals made by younger workers should be more in line with their supervisors' evaluations, resulting in higher interrater agreement. Atwater and Yammarino (1997) further speculated that longer-tenured employees may give less credence to work-related feedback provided by their bosses. This reluctance to listen to constructive feedback may reflect either overconfidence or complacency on the part of the more experienced worker, which is then reflected in higher self-ratings of performance, ultimately yielding lower rating agreement between these parties.

Supervisors may also be disinclined to give honest performance feedback to older employees, worrying that critical comments could negatively affect a long-standing relationship. In sum, either the absence of honest supervisory feedback or the disregarding of this feedback may result in inflated self-ratings for these longer-term employees which may, in turn, lessen the likelihood of their self-ratings correlating with supervisory ratings.

Applying Atwater and Yammarino's (1997) reasoning to a different rating pair (i.e., self- and peer raters), we expect more lenient self-ratings in longer-term coworker relationships, which may then account for decreased rating agreement in such relationships. Accordingly, the following hypothesis is offered:

Hypothesis 2: The target employee's self-ratings will be higher in longer-term peer relationships than in shorter-term relationships.

Trustworthiness of the Self-Rater

Previous research has shown that self-ratings made by more able employees agree more with the ratings of others when compared with self-ratings made by less able employees (Atwater & Yammarino, 1997). Self-raters with less ability seem less receptive to feedback on their performance, whereas self-raters with higher ability appear to be more open to learning about themselves and discovering strategies for improving future performance (Ruble & Flett, 1988). In their appraisal study, Furnham and Stringfield (1994) used the supervisor's performance evaluations instead of the more focused employee-ability measure and found that the target employee's self-ratings and subordinate's ratings tended converge more when the supervisor's appraisal ratings of the target employee were higher. They speculated that more effective workers tend to have greater insight into their own managerial behavior, which then yields higher self-rating accuracy and higher interrater agreement. Apparently, personal insight coupled with a willingness to receive feedback helps the highly able worker to make more veridical self-ratings than the less able worker (Furnham & Stringfield, 1994).

In attempt to extend the literature beyond the moderating effects of employee ability and overall performance-appraisal scores, we plan to test whether another construct, McAllister's (1995) cognition-based trust, can operate as a moderator of rating agreement. McAllister's (1995) cognition-based trust assesses the target employee's trustworthiness in workplace matters (e.g., dedication, professionalism, carefulness, dependability, competence). We propose then that employees who are characterized as more trustworthy by their coworkers (i.e., higher in cognition-based trust) will tend to be more careful and serious self-raters. That is, making trustworthy self-ratings is a likely subset of trustworthiness in one's general workplace affairs. In line with this thinking, we propose that:

Hypothesis 3: Self-ratings made by more trustworthy employees (i.e., higher in cognition-based trust) will agree with coworker ratings to a much greater extent than will the self-ratings made by less trustworthy employees.

Considering that employee appraisal is being touted in Russian firms (Kiriazov et al., 2000) and more indirectly in Polish firms (Sood & Mroczkowski, 1994), empirical research in this area deserves attention. Not surprisingly, self-rating and rating-agreement research in firms located in Russia and Poland has been given little attention thus far. After searching the international literature, we did not find any self-rating or rating-agreement studies involving Russian subjects. However, one Polish self-rating study was located. In that study, Schultz (2001) reported the tendency for some Polish self-raters employed at a chemical company to rate their own performance lower than their supervisors. Given the paucity in this line of research, we do not plan to offer formal hypotheses that suggest differences in interrater agreement between our Russian, Polish, and U.S. samples. When testing our three hypotheses, any differences detected in rating agreement between countries will be treated as exploratory, providing a preliminary look into the self- and peer-rating processes in these countries.

METHOD

Participants

There are 560 participants in this study, organized into 280 rating dyads. The Russian sample is made up of target employees from different firms located in and around city of Krasnoyarsk in central Russia. The target employee-coworker dyads in the Russian sample numbered 86. The Russian participants were recruited from among current and former students of a local university, who were employed in various firms. The Polish sample is made up of 99 employee-coworker dyads from different firms located in Warsaw. These Polish employees were enrolled in a non-traditional program for adult students in a major Polish university while working in their respective organizations. The U.S. sample is made up of 95 dyads, consisting of 95 target employees and 95 coworkers from mostly different firms located in Western New York State. Eighty-eight U.S. participants were employed in their respective organizations while enrolled as part-time students in adult evening programs at a local university. Another seven employees worked full-time in a small U.S. technology firm. In sum, behavioral ratings made by 280 dyads employed in different firms in three countries were included in this study. Demographic statistics describing each country's sample are provided in Table 1. As shown in Table 1, the average tenure of the U.S., Russian, and Polish participants in their firms is 7.49, 9.29, and 3.91 years, respectively, suggesting relatively experienced employees in this study.

TABLE 1

Descriptive Statistics for the Sample in Each Country

	<u>U.S.</u>	<u>Russia</u>	<u>Poland</u>
Target Employees:			
Female	56%	62%	70%
Management Position	38%	34%	18%
Average number of years in firm	7.49 yrs. (SD = 6.26)	9.29 yrs. (SD = 9.07)	3.91 yrs. (SD = 4.72)
Coworkers:			
Female	55%	70%	64%
Average number of years that coworker has known target employee	4.73 yrs. (SD = 4.92)	6.12 yrs. (SD = 6.38)	4.36 yrs. (SD = 12.10)
Average number of employees in firm	10,778 (SD = 23549)	895 (SD = 1859)	2030 (SD = 12783)

Measures

The questionnaire administered to target employees and coworkers included various demographic items (e.g., gender, number of years that they had known each other) and five items assessing the target employee's performance in various behavioral areas. Both sets of ratings were focused on the target employee's level of workplace behavior. Specifically, target employees were asked to evaluate their own performance in five areas: creativity, motivation, risk taking, initiative, and assertiveness. The same behavioral items were included, with appropriate minor changes in wording, in the coworker's questionnaire. The exact wording of the five behavioral items in the two questionnaires are: "My (his/her) level of creativity in the job is generally very high," "I (s/he) display(s) a high level of motivation and energy in my (the) job," "I (s/he) am not afraid of taking risks and trying new things in my (the) job," "I (s/he) take immediate action to resolve work-related problems as they emerge," and "I (s/he) volunteer my (his/her) work-related views without waiting to be asked." A five-point Likert scale format (1 = "strongly agree," ..., 5 = "strongly disagree") was used as anchors in the previously mentioned behavioral scales. After finding similar levels of internal consistency for the target employee's five behavioral ratings and for the coworker's five behavioral ratings for each country's sample (see the Cronbach alphas in Table 2), composite means were computed. These composites (i.e.,

TABLE 2
Means, Standard Deviations, and Intercorrelations^a

Variable	Mean	SD	1	2
United States: ^b				
1. Self-rating composite	4.08	.49	(.69)	
2. Coworker's composite	4.08	.72	.16	(.87)
Poland: ^c				
1. Self-rating composite	3.63	.60	(.71)	
2. Coworker's composite	3.71	.69	.11	(.79)
Russia: ^d				
1. Self-rating composite	3.80	.49	(.75)	
2. Coworker's composite	3.85	.69	.21*	(.84)

^a Cronbach's alphas appear on the diagonal for multiple-item measures.

^b $n = 95$.

^c $n = 99$.

^d $n = 86$.

* $p < .05$ (two-tailed test). ** $p < .01$ (two-tailed test).

a composite mean of the self-rater's five behavior ratings and a composite mean of the coworker's five ratings of the target employee) then served as the key rating measures in this study. The coworker's questionnaire also included McAllister's (1995) cognition-based trust measures (e.g., "This person approaches his/her job with professionalism and dedication," "Given this person's track record, I see no reason to doubt his/her competence and preparation for the job," "I can rely on this person not to make my job more difficult by careless work") in order to assess the target employee's overall trustworthiness. The same five-point Likert scale was used as anchors in this cognition-based trust scale. After finding an acceptable level of internal consistency for this multiple-item trustworthiness scale for each country (i.e., Cronbach alpha of .83 for the U.S., .79 for Russia, and .80 for Poland), a composite mean of these coworker trustworthiness ratings was computed. To avoid confusion when interpreting data, all ratings were reverse-scored. After reverse-scoring, a five score would be the highest level and a one would be the lowest. The Polish and Russian questionnaire items and rating instructions

were translated into their respective language. In line with Pavett and Morris (1995), all measures and instructions were then back-translated into English by a different interpreter in order to insure translation accuracy.

Procedures

Target employees in this study completed their self-ratings and then returned their responses in person or by mail. Similar to McAllister's (1995) procedures, these employees were also asked to deliver a similar questionnaire to a coworker of their choice. In the instructions, the coworker was told that they could return their completed ratings and demographic responses by mail or have their fellow employee return their sealed responses. All questionnaires were completed anonymously. An identifying number was used to match each target employee's self-ratings with his or her coworker's ratings of the target employee. The estimated response rate for the U.S. sample is 35% and 50% for both the Polish and Russian samples. All participants were instructed that their questionnaire responses were to be used for research purposes. That is, the researchers were the only individuals seeing all of the participants' ratings in this study. This rating-purpose instruction is consistent with the instructions given in other international self-rating studies (Fahr, Dobbins, & Cheng, 1991; Furnham & Stringfield, 1994; Yu & Murphy, 1993) and U.S.-based self-rating studies (Brief et al., 1977; Shore & Bleicken, 1991).

RESULTS

Besides internal consistency data, Table 2 presents descriptive statistics and intercorrelations between the self-rating composite and coworker composite for each country. A hierarchical regression was employed to test Hypotheses 1 and 3. Specifically, multiple regressions were used to assess the variance accounted for by the main predictor variable (i.e., coworker's composite rating) and related interactions on the dependent variable (i.e., the self-rater's composite rating). Our approach of using multiple regression to assess rating agreement between two parties parallels the statistical procedures used in other rating-agreement studies (Lee & Son, 1999; Shore & Bleichen, 1991). Because the national country variable is categorical, it was formed into two dummy variables. The coding for the first of these variables is: U.S. = 0, Poland = 1, and Russia = 0. The coding for the second dummy variable is: U.S. = 0, Poland = 0, and Russia = 1. The first dummy variable contrasts Poland with the U.S. and Russia, and the second dummy variable contrasts Russia with the U.S. and Poland (Aiken & West, 1991). Dummy variables were also formed for the target employee's gender (males = 0 and females = 1) and the coworker's gender (males = 0 and females = 1).

Hypothesis 1 proposed that rating agreement between the target employee's self-rating composite and the coworker's composite would be higher, the shorter the time that the two employees had known each other. To test Hypothesis 1, the hierarchical regression procedure consisted of six steps. In line with Oldham and Cummings (1996), all main effects in this study were entered into the regression equation in the first two steps, the two-way interactions followed in Steps 3, 4, and 5. The three-way interactions followed in Step 6. Specifically, to control for the influence of demographic differences between the three countries' samples, two demographic variables (i.e., both gender dummy variables) entered with the two country dummy variables and the number-of-years variable (i.e., number of years that the coworker had known

the target employee) in the regression equation in Step 1. The coworker's composite rating entered into the equation in Step 2. To determine whether the number-of-years variable moderates the relationship of the coworker's composite rating to the dependent variable (i.e., self-rating composite), this two-way interaction (number of years X coworker composite rating) entered in Step 3. The two two-way interactions involving the coworker's composite rating and the two country dummy variables entered in Step 4. The number-of-years X two country dummy variables interactions entered in Step 5. The two three-way interactions (number of years X coworker's composite rating X country dummy variables) entered in Step 6.

Due to multicollinearity that is present when both a main effect and interaction are included in the same regression model, which can result in unstable and non-meaningful regression coefficients, we judged the strength of the coworker composite variable and hypothesized interactions by the increase in variance accounted for (i.e., increase in R^2), when that variable was added to the regression model (see Oldham & Cummings, 1996).

As shown in Table 3, the hierarchical regression results suggest preliminary support for this first hypothesis. As expected, the coworker's composite rating accounted for significant incremental variance in the self-rating composite ($\Delta R^2 = .03$, $p < .01$) in Step 2. The coworker's composite rating X number-of-years interaction also accounted for significant additional variance ($\Delta R^2 = .04$, $p = .01$) in Step 3 of the regression, suggesting that the effect of the coworker's composite rating on the target employee's self-rating composite was moderated by the number of years that the coworker had known the target employee. Finally, the two three-way interactions in Step 6 (number of years X coworker's composite rating X country dummy variables) were also significant ($\Delta R^2 = .03$, $p < .05$), suggesting that the effect of the coworker's composite rating on the self-rating composite was not only moderated by the number of years that the coworker had known the target employee but also by country status. To determine the exact nature of this three-way interaction in Step 6 which subsumes the significant two-way interaction in Step 3, we followed the statistical procedure recommended by Oldham and Cummings (1996). Two different follow-up regression procedures were utilized to determine which country's level of rating agreement differed from the other two. Specifically, the first of the two follow-up regressions incorporated the same predictor variables as before, except that only one country dummy variable (U.S. versus Russia) and related interaction terms were included. The second follow-up regression incorporated the same variables as the previous one, except that the U.S.-Polish dummy variable and related interaction terms were included. The U.S./Russian regression results again showed a significant three-way interaction in Step 6, ($\Delta R^2 = .03$, $p < .05$), whereas the U.S./Polish regression results indicated that the only significant interaction was in Step 3 (number of years X coworker's composite rating, $\Delta R^2 = .07$, $p < .001$). Hence, the relationship between the self-rating composite and the coworker's composite rating appears to be moderated by both number of years and country status. Similar to Oldham and Cummings' (1996) procedures, the number-of-years variable was then split into two levels (years below each country's median versus years above the median) to determine more precisely how these two variables moderate rating agreement. The median number of years is 3.00, 1.21, and 4.00 for the U.S., Polish, and Russian samples, respectively. The correlations between the coworkers composite rating and the self-rating composite in the lower number-of-years

TABLE 3

**Summary of Hierarchical Regression Results of Number of Years and
Coworker Composite on Self-Rating Composite**

Variable	Self-Rating Composite	
	R ²	Δ R ²
Step 1: Demographic Variables: Gender, number of years, and country dummy variables	.08	.08***
Step 2: Coworker rating composite	.11	.03**
Step 3: Number of years X Coworker composite	.15	.04**
Step 4: Country dummy variables X Coworker composite	.15	.00
Step 5: Number of years X Country dummy variables	.16	.01
Step 6: Number of years X Country dummy variables X Coworker composite	.19	.03*

n = 280.

p* < .05. *p* < .01. ****p* < .001.

condition (i.e., lower than the country's median) are $r = .31$ ($p < .05$), $r = .36$ ($p < .01$), and $r = .15$ ($p = .32$) for the U.S., Polish, and Russian samples, respectively. The correlations between the coworkers composite rating and the self-rating composite in the higher number-of-years condition (i.e., higher than the country's median) are $r = .00$ ($p = .99$), $r = .04$ ($p = .77$), and $r = .20$ ($p = .21$) for the U.S., Polish, and Russian samples, respectively. In sum, these higher correlations suggest greater rating agreement between the target employee and coworker in relatively new relationships for the U.S. and Polish samples only, whereas the insignificant Russian correlations in both number-of-years conditions suggest that rating agreement between Russian raters was consistently weak in both shorter-term and longer-term relationships. Taken as a whole, these results suggest support for Hypothesis 1 for the U.S. and Polish samples, but not the Russian sample.

Hypothesis 2 proposed that the target employee's self-rating composite will be higher in longer-term peer relationships. To test this hypothesis, we again utilized the dichotomized number-of-years variable as an independent variable (i.e., below the median is the low level and above the median is the high level). We conducted a 3 x 2 (U.S./Poland/Russia X low/high number of years in the relationship) analysis of covariance (ANCOVA) to test whether the level

of the self-rating composite differed by number of years in the relationship and by country. The composite self-rating served as the dependent measure in this study. To control for possible extraneous effects of gender differences between the three national samples on the composite ratings, two demographic variables (i.e., target employee's gender and coworker's gender) served as covariates in this analysis. The means for each country's self-rating composite are presented in Table 4. The ANCOVA results indicated that the number-of-years main effect, $F(1,260) = .16, p = .69$, and the two-way number-of-years X country interaction, $F(2, 260) = 1.31, p = .27$, are not significant, suggesting that the length of time in the relationship did not produce the expected higher self-ratings in longer-term relationships across the three countries. These findings do not support Hypothesis 2.

TABLE 4

Mean Self-Rating Composite by Number-of-Years in Relationship

Country	Self-Rating Composite	
	M	SD
United States (N = 92):		
Shorter-term relationship (below median)	4.12	.47
Longer-term relationship (above median)	4.05	.53
Russia (N = 86):		
Shorter-term relationship (below median)	3.73	.51
Longer-term relationship (above median)	3.85	.44
Poland (N = 99):		
Shorter-term relationship (below median)	3.68	.55
Longer-term relationship (above median)	3.55	.66

* $p < .05$ level. ** $p < .01$ level.

Hypothesis 3 proposed that the self-ratings made by more trustworthy employees (i.e., higher in cognition-based trust) will agree with coworker ratings to a much greater extent than

will the self-ratings made by less trustworthy employees. To test this hypothesis, we first formed the employee-trustworthiness variable into a dichotomized moderator variable. Each country's median for the coworker cognition-based trust variable was used to form two levels (i.e., below the median is a lower level of trustworthiness [dummy-code = 0], and above the median is a higher level of trustworthiness [dummy code = 1]). The median trustworthiness rating is 4.40, 4.00, and 4.00 for the U.S., Russian, and Polish samples, respectively. Such high median trust scores are not surprising considering that these ratings were completed by a coworker chosen by the target employee. Except for this new trustworthiness variable, which replaces the number-of-years variable in the previous regression equation, the exact same demographic variables and predictor variables used to test Hypothesis 1 were utilized to test Hypothesis 3. These variables were entered into the regression equation in the same order as before. As shown in Table 5, the regression results suggest preliminary support for the third hypothesis. The coworker's composite rating accounted for significant incremental variance ($\Delta R^2 = .03, p < .01$) in Step 2.

TABLE 5

**Summary of Hierarchical Regression Results of Cognition-Based Trust
and Coworker Composite on Self-Rating Composite**

Variable	Self-Rating Composite R^2	ΔR^2
Step 1: Cognition-based trust and demographic variables: gender and country dummy variables	.08	.08***
Step 2: Coworker rating composite	.11	.03**
Step 3: Cognition trust X Coworker composite	.13	.02*
Step 4: Country dummy variables X Coworker composite	.13	.01
Step 5: Cognition trust X Country dummy variables	.14	.01
Step 6: Cognition trust X Country dummy variables X Coworker composite	.15	.00

$n = 280.$

** $p < .05.$ ** $p < .01.$ *** $p < .001.$

The coworker's composite rating X target-employee trustworthiness interaction also accounted for significant additional variance ($\Delta R^2 = .02$, $p < .05$) in Step 3 of the regression, suggesting that the effects of the coworker's composite rating on the target employee's self-rating composite are moderated by the perceived trustworthiness of the target employee. Finally, the two three-way interactions in Step 6 (trustworthiness X coworker's composite rating X country dummy variables) are not significant ($\Delta R^2 = .00$, $p = .50$), suggesting that the effects of the coworker's composite rating on the self-rating composite is not moderated by both the target employee's trustworthiness and by country status. To determine the exact nature of the previously reported two-way interaction (coworker's composite rating X trustworthiness) in Step 3, we followed Oldham and Cummings' (1996) procedure and employed another employee-trustworthiness dummy variable which was formed around the cognition-based trust median (i.e., 4.00) for the whole sample (i.e., across the three countries). Specifically, a cognition-based trust composite score of 4.00 or above was dummy-coded 1, which suggests that the peer rater had little-to-no reservation on the target employee's overall trustworthiness. Conversely, a cognition-based trust score of 3.99 or lower was dummy-coded 0, which suggests that the peer rater had more reservation on the target employee's trustworthiness. We then looked at the correlation between the coworker composite rating and the self-rating composite in each employee-trustworthiness condition. These findings reveal a significant relationship between the self-rating composite and coworker's composite in the high-trustworthiness condition ($r = .27$, $p < .01$), whereas the correlation between these composites is not significant in the low-trustworthiness condition ($r = .13$, $p = .11$). These results suggest greater agreement between the self-rater and coworker when the target employee is considered highly trustworthy versus less trustworthy. Taken as a whole, these results suggest support for Hypothesis 3 for all three countries.

DISCUSSION

Russian researchers (e.g., Kiriazov et al., 2000) have advocated some form of performance assessment in Russian firms. Similarly, Polish researchers such as Sood and Mroczkowski (1994) have stated that the biggest challenge facing Polish firms is the poor work ethic, signifying that individual accountability and measurement may be needed. Identifying appraisal processes that are better suited to Russia and Poland may help with the transformation of the old work ethic to an improved one. One feedback process being considered in Eastern European organizations is multi-source feedback systems (Kiriazov et al., 2000). Whether such a feedback system will transfer to the Eastern European organizational setting is, according to Kiriazov et al. (2000), unsettled. This study attempts to join this debate by presenting empirical findings on one aspect of the multi-source feedback process, rating agreement among Russian and Polish self- and peer raters.

Historically, self- and peer ratings have been found to be weakly correlated across studies (Conway & Huffcutt, 1997). Discounting the moderator effects detected in this study, our Russian, Polish, and U.S. self- and peer raters fared no better in producing rating agreement (see Table 2) than did the 6,359 self- and peer raters included in Conway and Huffcutt's (1997) meta-analysis. Are self- and peer ratings destined to such low interrater agreement across organizational settings and countries or might an organizational intervention, such as frame-of-reference training (Bernardin, Buckley, Tyler, & Wiese, 2000), raise interrater agreement between these rating sources to a more respectable level? Future research should probe more

deeply into these questions.

Our search for moderators was intended to offer some insight into the self- and peer-rating processes. The moderator, time in a coworker relationship, had no effect on our Russian results. The Polish and U.S. self- and peer raters, however, agreed more when they had known each other a shorter period of time. Such a finding seems counterintuitive because one would think that more observation time translates into higher quality ratings (e.g., increased interrater agreement). Uncovering the underlying reason for this U.S. and Polish finding is therefore important if interrater agreement research is to move forward. Unfortunately, our ANCOVA results do not support the argument posed in Hypothesis 2 that suggests that distorted, inflated self-ratings in longer-term relationships accounts for the lower interrater agreement in this condition. An alternative explanation should be explored in future studies.

In this regard, future research should consider whether the content and frequency of the communication between coworkers change, the longer the relationship lasts. Similar to the communication that occurs in more intimate personal relationships, perhaps the amount of interpersonal sharing on work-related concerns between coworkers lessens over time, resulting in less understanding of the other party's job-related difficulties and failures. In other words, rating consensus between coworkers may be easier to reach with regular sharing of individual concerns about work-related responsibilities and outcomes. The absence of on-going sharing may reduce the likelihood of attaining interrater agreement. Future research should consider testing this notion utilizing a within-subjects experimental design. In this way, changes in coworker communication processes and corresponding self- and peer ratings could be tracked over the life of each relationship.

The results of this study also provide evidence for another moderator of rating agreement which is employee trustworthiness (i.e., cognition-based trust). Our findings speak to the importance of having trustworthy employees, who are generally dedicated, professional, and careful in their work assignments, involved in multi-source feedback programs. Even though the median cutoff for more versus less trustworthiness was quite high in the present study, we still detected the moderating effects of this variable across the three national samples. These findings suggest that target-employees perceived to be lower in cognition-based trust may not only be less trustworthy in their everyday job matters but also less dependable and less professional when making self-ratings. Whether education and training can mitigate the undesirable rating effects originating from a less responsible work ethic may be worth exploring.

Assuming that these results generalize to other U.S., Polish, and Russian samples, it seems that cross-cultural value differences (e.g., Hofstede, 1980) did not influence interrater agreement in the present study. As reported in Hofstede (2001), Russia and Poland's indices on Hofstede's dimensions (e.g., power distance, collectivism-individualism) markedly differ from the U.S. However, the moderating effects detected in this study show a consistent pattern for the U.S. and Polish raters and a less consistent pattern for the Russian raters. If cultural differences were influencing our study's ratings, one would instead expect greater consistency for the Russian and Polish ratings when contrasted with the U.S. ratings.

Some caution should be expressed in that there is always the possibility that demographic differences in the three samples may have influenced this study's findings. That is, the median statistic used to form the time-in-relationship moderating variable varies between the three national subgroups. Would we find the same results if each national sample's level of "time in the relationship" was more consistent between countries? Our results should therefore be viewed as preliminary, acting as a catalyst for further multi-source feedback research in Eastern

European organizations. The present study's results also could be replicated with demographically diverse samples in other Polish and Russian locations.

Our sampling strategy, which is similar to McAllister's (1995) procedures, may raise some concern on the generalizability of our findings. Because most of the participants in our study were directly or indirectly recruited from a university setting, our samples may have an unrepresentative number of younger, educated employees. It bears repeating that the Russian, Polish, and U.S. participants in the present study were attending or associated with adult, non-traditional education programs while working for their respective firms. The average tenure of the U.S., Russian, and Polish participants in their organizations was 7.49, 9.29, and 3.91 years, respectively. These employees should not be viewed as inexperienced in the work setting. Nonetheless, our samples may not be representative of each country's workforce.

REFERENCES

- Atwater, L.E., & Yammarino, F.J. (1997). Self-other rating agreement: A review and model. *Research in Personnel and Human Resources Management*, 15, 121-174.
- Bernardin, H.J., Buckley, M.R., Tyler, C.L., & Wiese, D.S. (2000). A reconsideration of strategies in rater training. *Research in Personnel and Human Resources Management*, 18, 221-274.
- Brief, A.P., Aldag, R.J., & VanSell, M. (1977). Moderators of the relationships between self and superior evaluations of job performance. *Journal of Occupational Psychology*, 50(2), 129-134.
- Brutus, S., Fleener, J.W., & McCauley, C.D. (1999). Demographic and personality predictors of congruence in multi-source ratings. *The Journal of Management Development*, 18(5), 417-433.
- Bush, J. (2003). Sizzling growth could singe Russia's economy. *Business Week*, (Issue 3843) (July 28, 2003), 52.
- Condon, C., & Butler, R. (2003). A chill east wind Eastern Europe's dynamos are losing jobs to Asia. *Business Week*, (Issue 3847), 44.
- Conway, J.M., & Huffcutt, A.I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance*, 10(4), 331-360.
- Farh, J.L., Dobbins, G.H., & Cheng, B.S. (1991). Cultural relativity in action: A comparison of self-ratings made by Chinese and U.S. workers. *Personnel Psychology*, 44(1), 129-147.
- Fey, C., Engstrom, P., & Bjorkman, I. (1999). Effective human resource management practices for foreign firms in Russia. *Organizational Dynamics*, 28(2), 69-80.
- Fletcher, C., & Baldry, C. (2000). A study of individual differences and self-awareness in the context of multi-source feedback. *Journal of Occupational and Organizational Psychology*, 73(3), 303-319.
- Furnham, A., & Stringfield, P. (1994). Congruence of self and subordinate ratings of managerial practices as a correlate of supervisor evaluation. *Journal of Occupational and Organizational Psychology*, 67(1), 57-67.
- Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Newbury Park, CA: Sage.

- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks, CA: Sage.
- Kiriazov, D., Sullivan, S.E., & Tu, H.S. (2000). Business success in Eastern Europe: understanding and customizing HRM. *Business Horizons*, 43(1), 39-43.
- Kozminski, A.K. (1999). *Zarządzanie międzynarodowe. Konkurencja w klasie światowej*. Warszawa, Poland: PWE.
- Lee, M., & Son, B. (1999). The agreement between self- and supervisor ratings: An investigation of leader-member exchange effects. *International Journal of Management*, 16(1), 77-88.
- McAllister, D.J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, 38(1), 24-59.
- Mersman, J.L., & Donaldson, S.I. (2000). Factors affecting the convergence of self-peer ratings on contextual and task performance. *Human Performance*, 13(3), 299-322.
- Oldham, G.R., & Cummings, A. (1996). Employee creativity: Personnel and contextual factors at work. *Academy of Management Journal*, 39(3), 607-634.
- Pavett, C., & Morris, T. (1995). Management styles within a multinational corporation: A five-country comparative study. *Human Relations*, 48(10), 1171-1191.
- Ruble, D., & Fleet, G. (1988). Conflicting goals in self-evaluative information seeking: Developmental and ability level analysis. *Child Development*, 59(1), 97-106.
- Schultz, A. (2001). Ocena i samoocena kadry kierowniczej. *Humanizacja Pracy*, 199-200, 59-66.
- Shore, L.M., & Bleicken, L.M. (1991). Effects of supervisor age and subordinate age on rating congruence. *Human Relations*, Vol. 44(10), 1093-1105.
- Sood, J., & Mroczkowski, T. (1994). Human resource management challenges in Polish private enterprise. *International Studies of Management and Organization*, 24(4), 48-63.
- Warr, P., & Bourne, A. (1999). Factors influencing two types of congruence in multirater judgments. *Human Performance*, 12(3/4), 183-210.
- Yammarino, F.J., & Atwater, L.E. (1997). Implications of self-other rating agreement for human resources management. *Organizational Dynamics*, 25(4), 35-44.
- Yu, J., & Murphy, K.R. (1993). Modesty bias in self-ratings of performance: A test of the cultural relativity hypothesis. *Personnel Psychology*, 46(2), 357-363

Robert D. Costigan (costigan@sjfc.edu) is an Associate Professor of Management, St. John Fisher College, 3690 East Avenue, Rochester, NY 14618.

Richard C. Insinga (insingrc@oneonta.edu) is an Associate Professor of Management, State University of New York, Ravine Parkway, 226 Netzer, Oneonta, NY 13820.

Grazyna Kranas (gkranas@sci.psych.uw.edu.pl) is a Professor of Psychology, University of Warsaw, ul. Stawki 5/7, 00-183 Warsaw, Poland.

Selim S. Ilter (ilter@sjfc.edu) is a Professor of Management, St. John Fisher College, 3690 East Avenue, Rochester, NY 14618.

Vladimir A. Kureshov (kureshov@stu.krasnoyarsk.su) is a Vice Rector for International Cooperation, Siberian State Aerospace University, 31 Krasnoyarsky Robochy, Krasnoyarsk, 66014, Russia.

J. Jason Berman (berman@sjfc.edu) is a Professor of Management, St. John Fisher College, 3690 East Avenue, Rochester, NY 14618.