12-2010

Measurement equivalence of the Wong and Law Emotional Intelligence Scale across self and other ratings

Nele LIBBRECHT
Ghent University

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg

Eveline SCHOLLAERT
Ghent University

DOI: https://doi.org/10.1177/0013164410378090

Educational and Psychological Measurement
http://epm.sagepub.com/

Measurement Equivalence of the Wong and Law Emotional Intelligence Scale Across Self and Other Ratings

Nele Libbrecht, Filip Lievens and Eveline Schollaert
Educational and Psychological Measurement 2010 70: 1007
DOI: 10.1177/0013164410378090

Educational and Psychological Measurement
70(6) 1007–1020

DOI: 10.1177/0013164410378090
http://epm.sagepub.com
SAGE

Measurement Equivalence of the Wong and Law Emotional Intelligence Scale Across Self and Other Ratings

Nele Libbrecht[1], Filip Lievens[1], and Eveline Schollaert[1]

Abstract

There exist a variety of measurement instruments for assessing emotional intelligence (EI). One approach is the use of other reports wherein knowledgeable informants indicate how well the scale items describe the assessed person's behavior. In other reports, the same EI scales are typically used as in self-reports. However, it is not known whether the measurement structure underlying EI ratings is equivalent across self and other ratings. In this study, the measurement equivalence of an extant EI measure (Wong and Law Emotional Intelligence Scale [WLEIS]) across self and other ratings was tested. Using multiple group confirmatory factor analysis, the authors conducted a sequence of increasingly more restrictive tests of equivalence across self and other ratings. The WLEIS was found to be configurally and metrically invariant across self and other ratings. However, there was no evidence for structural invariance between rater groups. Future research is needed to test the equivalence of other EI measures across self and other ratings.

Keywords

emotional intelligence, measurement invariance, confirmatory factor analysis, emotional intelligence assessment

Since the mid-1990s, emotional intelligence (EI) has been studied extensively and today there are a variety of measurement instruments available for assessing EI. An important distinction concerns the method of measurement (performance-based vs.

[1]Ghent University, Ghent, Belgium

Corresponding Author:
Nele Libbrecht, Department of Personnel Management, Work and Organizational Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium
Email: nele.libbrecht@ugent.be

self-report). In performance-based EI tests, people are presented with emotion-based problem-solving items and the quality of their answers is evaluated using predetermined criteria (Beaupré, Cheung, & Hess, 2000; Freudenthaler & Neubauer, 2005; Mayer, Salovey, & Caruso, 2002). Conversely, in self-report EI measures, individuals are asked to report their own level of EI, as respondents are presented with descriptive statements and indicate the extent to which they agree or disagree with the statements (Brackett, Rivers, Shiffman, Lerner, & Salovey, 2006; Pérez, Petrides, & Furnham, 2005; Schutte et al., 1998; Wong & Law, 2002). Although self-report EI questionnaires are often used to assess EI, they have some key drawbacks. One problem is that self-report EI measures are likely to reflect perceived rather than actual performance levels. Previous research has shown that there are only small correlations between perceived estimates of ability and actual abilities (Davies, Stankov, & Roberts, 1998; Paulhus, Lysy, & Yik, 1998). For instance, Paulhus et al. (1998) found rather low correlations between self-reports of intelligence and IQ scores (.20-.25), and Brackett et al. (2006) found that about 80% of people believe that they are among the 50% most emotionally intelligent people. Another problem is that self-report EI measures can easily be faked good (Day & Carroll, 2007; Grubb & McDaniel, 2007).

In light of the drawbacks of self-report EI measures, other reports have been suggested as an alternative approach for measuring EI (Charbonneau & Nicol, 2002; Jordan, Ashkanasy, Hartel, & Hooper, 2002; Law, Wong, & Song, 2004; Lopes, Grewal, Kadis, Gall, & Salovey, 2006; Song et al., 2010; Van der Zee, Thijs, & Schakel, 2002). In other reports, knowledgeable informants such as friends and colleagues indicate how well the scale items describe the focal person's behavior. It can be argued that information that is relevant for judging traits and competencies is better perceived by others than by the target person him or herself. It can also be expected that the measurement provided by others is less lenient and more reliable (Van der Zee et al., 2002). Another reason to use peers for measuring EI is that the interpersonal component of EI can be assessed with other reports. Finally, other reports share the self-reports' advantage of being inexpensive and easy to administer in a larger groups of respondents. Given these advantages, other ratings might provide a valuable alternative approach for the assessment of EI.

In other reports, the same EI scales are typically used as in self-reports. For example, the Emotional Competence Inventory (Boyatzis, Goleman, & Rhee, 2000) includes 360-degree assessment techniques that can be completed by peers and supervisors. In addition, the Workgroup Emotional Intelligence Profile (WEIP; Jordan et al., 2002) measures EI of individuals who work in teams and contains both a self-reporting and a peer-reporting measure. Another example is the Wong and Law Emotional Intelligence Scale (WLEIS, Wong & Law, 2002) that can be rated by parents, supervisors, and peers (Law et al., 2004; Song et al., 2010).

Thus, a key assumption of current EI measurement practice is that the same (originally developed as self-report) EI questionnaires can be used for both self and other ratings. However, it is not known whether the measurement structure underlying EI ratings is equivalent across self and other ratings. For example, it might well be that intrapersonal EI dimensions (e.g., the ability to understand and express one's own emotions or to effectively use one's emotions) have a limited observability for others. In light of this difference in observability, self and other raters might use a different conceptual frame-of-reference when rating items of an EI questionnaire, leading to lack of measurement invariance. Yet measurement equivalence of any questionnaire across different rating groups is a prerequisite for making substantive interpretations of the ratings and formulating accurate recommendations. Therefore, the assumption that the measurement instrument means the same to different rater groups and functions the same across these rater groups must be tested (Cheung & Rensvold, 2002; Vandenberg & Lance, 2000).

The purpose of this study is to test the measurement equivalence of an extant EI measure across self and other ratings. Using multiple group confirmatory factor analysis (MCFA), we conduct a sequence of increasingly more restrictive tests of equivalence across self and other ratings. Only if the EI test is measurement equivalent, this originally developed self-report EI measure can also be rated by others to assess an individual's level of EI.

Method

Participants and Procedure

A total of 341 final-year students (39% male, mean age = 23 years) voluntarily participated in sessions on psychological testing and assessment (i.e., simulated selection situation) that lasted 1 day. This simulated selection procedure gave them the unique opportunity to prepare themselves for applying for jobs in the labor market. The participants took the initiative to register for the sessions. These sessions were advertised across all majors of a large European University. There was anecdotal evidence that participants perceived the simulation to be comparable with real selection settings. For instance, they all reported to be anxious to take the tests.

At the start of the session, participants were told that they could increase their experience with taking a variety of tests. Next, they completed various tests, assessment center exercises, and questionnaires. Participants were also asked to distribute copies of questionnaires to two peers after the session. They were instructed to select friends that knew them well. In this study, we used only the scores on the EI test as rated by the focal person and their peers. Acquaintance between the focal person and the peers was measured with three items with a 5-point Likert-type scale (1 = *low acquaintance*; 5 = *high acquaintance*). If the average acquaintance score was less than 3, the peers were excluded from the study. Four participants were excluded from the study on the basis of their acquaintance scores.

We received ratings from at least one peer from 154 participants (45% of the 341 students who participated in the simulated selection procedure; 41% male; mean age = 23 years). These participants and their peer ratings were included in this study. When we received two peer ratings from a participant, we randomly selected a single peer rating for inclusion in the study. The second peer rating was then used to form a replication peer group. This replication sample (N = 115) consisted of one hundred fifteen participants and was used to examine whether the results obtained in the main sample could be replicated.

Wong and Law Emotional Intelligence Scale

The Wong and Law Emotional Intelligence Scale (WLEIS, Wong & Law, 2002) is a popular self-report measure of EI. As noted above, the WLEIS was designed to be used for self and other ratings. This EI scale is based on Davies et al.'s (1998) four-dimensional definition of EI. The WLEIS consists of 16 items with each subscale measured with 4 items. The Self Emotion Appraisal dimension assesses individuals' ability to understand and express their own emotions. A sample item is "I really understand what I feel." The Others' Emotion Appraisal dimension measures peoples' ability to perceive and understand the emotions of others. A sample item is "I always know my friends' emotions from their behavior." The Use of Emotion dimension denotes individuals' ability to use their emotions effectively by directing them toward constructive activities and personal performance. A sample item is "I always tell myself I am a competent person." The Regulation of Emotion dimension refers to individuals' ability to manage their own emotions. A sample item from this dimension is "I have good control of my own emotions." The WLEIS was measured with a 5-point Likert-type scale, ranging from 1 (*totally disagree*) to 5 (*totally agree*). Previous research has found support for the underlying four-factor structure, reliability, and convergent and discriminant validity of the WLEIS scores (Law et al., 2004; Law, Wong, Huang, & Li, 2008; Shi & Wang, 2007; Wong & Law, 2002). The WLEIS scores have also shown validity for predicting life satisfaction, academic performance, job performance, and job satisfaction (Song et al., 2010; Law et al., 2008; Wong & Law, 2002).

Analyses and Results

Test of Fit of Measurement Model Underlying the WLEIS (Within Each Group)

We began by testing several measurement models that represented different conceptualization underlying the WLEIS. First, we tested a one-factor model. This model represented the view of EI as a unitary construct. Second, we tested a two-factor model. This model specified EI as a multidimensional construct, consisting of a first factor representing the WLEIS subscales self appraisal, others' appraisal, and the use of emotions and a second factor representing the WLEIS subscale regulation of emotions. This two-factor model is based on the distinction in EI between a strategic area (i.e., emotional perception and facilitation of thought) and an experimental area (i.e., understanding and regulating emotions; Mayer, Salovey, Caruso, & Sitarenios, 2003). Finally, we tested a four-factor model wherein the four subscales of the WLEIS were hypothesized to represent four correlated latent factors and all items in a particular subscale were expected to load onto their designated factor. This model is consistent with the theoretical rationale underlying the WLEIS and has received the most empirical support in prior research (e.g., Law et al., 2004; Shi & Wang, 2007; Wong & Law, 2002).

To test the fit of these measurement models through confirmatory factor analysis (CFA) within each sample, we used EQS (Bentler, 1995). Several types of fit indices were used to assess how the CFA models represented the data. First, we used two absolute fit indices: χ^2 and the ratio of the χ^2 to its degrees of freedom (χ^2/df). Second, Bollen's incremental fit index (IFI) was used as relative fit index. Finally, two fit indices based on the noncentrality parameter were used, namely the comparative fit index (CFI), and the root mean square error of approximation (RMSEA). The criteria for determining whether the models give a good fit were for the χ^2 to be nonsignificant, the χ^2/df to be small approaching unity (Bentler, 1995), for the IFI and CFI to have values >.95, and for the RMSEA to be < .05.

Mardia's normalized multivariate kurtosis were 13.52 in the self group, 13.96 in the peer group, and 10.16 in the replication peer group. As the assumption of multivariate normality was violated, we used robust maximum likelihood estimation as estimation technique so that the χ^2 and standard errors were corrected for nonnormality (Satorra & Bentler, 1994).

Results of the CFAs by rater group are presented in Table 1. Only the four-factor model produced a good fit to the data in all groups, and was therefore used as baseline model in our further analyses. Note that the four-factor model fitted the data of the self rater group (Satorra–Bentler [S-B] χ^2 = 97.4) better than the data of the peer rater group (S-B χ^2 = 150.8) and the replication peer rater group (S-B χ^2 = 140.2).

Table 2 reports descriptive statistics and Cronbach's alpha for the four subscale scores of the WLEIS broken down by rater group. In addition, the associated 95% confidence intervals for the Cronbach's alphas are shown. With one exception, the alphas obtained in this study range between .74 and .90. These values exceed the .70 criterion deemed appropriate for instrument development research (see Henson, 2001).

Test of Invariance of the Measurement Model (Multiple Groups)

Once a baseline model was established within each rater group, we examined the invariance or equivalence of this model across both rater groups. To this end, we conducted a sequence of increasingly more restrictive tests of invariance across both rater groups via EQS (see Byrne, 1998; Hancock, 1997): (a) factor form (i.e., the same number of factors and the factors have the same variables loading on them), (b) factor pattern coefficients, (c) error variances, and (d) factor variances and covariances.

Table 1. Summary of Goodness-of-Fit Statistics for Within-Group Measurement Models of the Wong and Law Emotional Intelligence Scale.

	S-B χ^2	*df*	χ^2/df	IFI	CFI	RMSEA
Self rater group (*N* = 154)						
One-factor model	518.922	104	4.99	.41	.40	.16
Two-factor model	364.693	103	3.54	.63	.62	.13
Four-factor model	97.437	98	0.99	1.00	1.00	.00
Peer rater group (*N* = 154)						
One-factor model	849.568	104	8.17	.25	.24	.22
Two-factor model	577.213	103	5.60	.53	.52	.18
Four-factor model	150.823	98	1.54	.95	.95	.06
Replication peer rater group (*N* = 115)						
One-factor model	594.139	104	5.71	.40	.38	.21
Two-factor model	391.674	103	3.80	.64	.64	.16
Four-factor model	140.183	98	1.43	.95	.95	.06

Note: S-B χ^2 = Satorra–Bentler chi-square; *df* = degrees of freedom; IFI = Bollen's incremental fit index; CFI = comparative fit index; RMSEA = root mean square error of approximation.

A χ^2 difference test has typically been used to determine whether constraining parameters to be invariant across groups leads to a significant decrease in fit. However, this index has been criticized to be sensitive to sample size (Cheung & Rensvold, 2002). Hence, other indices of difference in fit have been proposed. Cheung and Rensvold (2002) showed that the CFI change was independent of model complexity and sample size. Based on simulations they determined that a ΔCFI value might not be higher than .01 to indicate that a constrained model did not lead to a meaningful drop of fit compared with an unconstrained model, and thus that the constrained parameters were invariant across groups.

Results of the sequence of the increasingly more restrictive tests of measurement invariance are presented in Table 3. In a first step, we tested for form invariance (also known as configural invariance) which implies that the model has no parameters constrained to be equal across both groups, except the number of factors and which variables loaded on the factors. The form invariance model yielded an adequate fit to the data, $\chi^2(196) = 249.192$ ($p < .00$), $\chi^2/df = 1.27$, IFI = .97, CFI = .97, and RMSEA = .04. Conceptually, this means that that self and peer raters use a similar frame of reference when completing the EI measurement scale (Riordan & Vandenberg, 1994).

In a second step, we constrained all factor pattern coefficients to be invariant across both rater groups. This test of measurement invariance is known as a test of metric invariance. As can be seen in Table 3, the goodness of fit related to the metric invariance model is adequate, $\chi^2(208) = 267.849$ ($p < .00$), $\chi^2/df = 1.29$, IFI = .97, CFI = .96, and RMSEA = .04. This additional set of constraints also did not lead to a meaningful drop in fit (ΔCFI = .004) as assessed by the guidelines of Cheung and Rensvold (2002). These results indicate that both rater groups calibrate the intervals used on the EI measurement scale in similar ways (Riordan & Vandenberg, 1994).

Table 2. Descriptive Statistics and Cronbach's Alphas of the Wong and Law Emotional Intelligence Scale Within Each Rater Group

Self rater group (*N* = 154)	Self Emotion Appraisal	Others' Emotion Appraisal	Use of Emotion	Regulation of Emotion
M	3.85	3.76	3.70	3.60
SD	0.60	0.57	0.59	0.70
Skewness	–0.32	–0.11	–0.16	–0.37
Kurtosis	0.97	0.09	0.43	–0.13
Cronbach's α	.79	.77	.76	0.82
95% CI of α	[0.73-0.84]	[0.71-0.83]	[0.70-0.82]	[0.77-0.86]
Peer rater group (*N* = 154)				
M	3.81	3.78	3.68	3.61
SD	0.67	0.66	0.64	0.76
Skewness	–0.90	–0.21	–0.35	–0.49
Kurtosis	1.53	–0.08	–0.36	–0.22
Cronbach's α	.86	.85	.74	.87
95% CI of α	[0.83-0.90]	[0.81-0.89]	[0.67-0.80]	[0.83-0.90]
Replication peer rater group (*N* = 115)				
M	3.93	3.80	3.71	3.65
SD	0.59	0.72	0.54	0.81
Skewness	–0.16	–0.61	0.01	–0.93
Kurtosis	–0.17	–0.10	–0.44	0.65
Cronbach's α	.77	.88	.65	.90
95% CI of α	[0.69-0.83]	[0.84-0.91]	[0.53-0.74]	[0.87-0.93]

Note: CI = confidence interval. The 95% CIs of Cronbach's alpha were computed on the basis of Fan and Thompson (2001).

In a third step, we tested for error invariance by constraining all error variances to be invariant across both rater groups. This model yielded a good fit to the data, $\chi^2(224) = 288.525$ ($p < .00$), $\chi^2/df = 1.29$, IFI = .96, CFI = .96, and RMSEA = .04. Adding this additional set of constraints did also not lead to a meaningful drop in fit (ΔCFI = .002).

Finally, we tested for structural invariance by constraining the invariance of factor variances and covariances across both rater groups. This model produced an adequate fit to the data, $\chi^2(234) = 314.335$ ($p < .00$), $\chi^2/df = 1.34$, IFI = .95, CFI = .95, and RMSEA = .05. However, although the ΔCFI value was not higher than .01 (ΔCFI = .009) some might argue that the addition of these parameters led to a meaningful decrease in fit. Inspection of the modification indices indeed showed that equal factor variance constraints did not hold for three of the four factor variances. More specifically, factor variances for the factors others' emotion appraisal, use of emotion, and regulation of emotion were higher for peer ratings than for self-ratings. This means that peer ratings showed a wider range in these WLEIS dimensions than the self-ratings (Cheung & Rensvold, 2002).

Table 3. Tests of Measurement Invariance for the Multigroup Measurement Model of the Wong and Law Emotional Intelligence Scale Across Self and Peer Raters

Model	S-B χ^2	*df*	IFI	CFI	RMSEA	Model Comparison	ΔS-B χ^2	Δdf	ΔCFI
1. Equal number of factors	249.192**	196	.969	.968	.043	—	—	—	—
2. Equal factor loadings	267.849**	208	.965	.964	.044	2 vs. 1	20.952	12	.004
3. Equal error variances	288.525**	224	.963	.962	.044	3 vs. 2	20.644	16	.002
4. Equal factor variances/ covariances	314.335**	234	.953	.953	.048	4 vs. 3	29.711	10	.009

Note: S-B χ^2 = Satorra–Bentler chi-square; IFI = Bollen's incremental fit index; CFI = comparative fit index; RMSEA = root mean square error of approximation.
**p* < .05. ***p* < .01.

Table 4 gives an overview of the parameters of the final model for the self and peer raters. Inspection of the parameter estimates confirms that the factor pattern coefficients and the factor covariances were invariant between both rater groups. However, as already stated, three of the four factor variances were noninvariant between both rater groups. Table 4 shows that factor variances for the factors others' emotion appraisal, use of emotion, and regulation of emotion were higher for peer ratings than for self-ratings.

To cross-validate these findings, we tested the same measurement invariance models in the replication peer sample. As can be seen in Table 5, the same pattern of results was found. So, we found evidence of configural and metric invariance. However, no evidence of structural invariance was established as the CFI change associated with the structural invariance test was higher than .01 (ΔCFI = .011). Note that we also conducted the measurement invariance tests with three groups, with the self raters as a first group, the peer raters as a second group, and the replication peer group as a third group. Again, the same results were found. Detailed results for the peer replication sample and three-group measurement invariance analyses are available from the first author.

Additional Analyses

We conducted additional analyses to examine the correlations between the latent factors of both rater groups. To this end, we tested an eight-factor model wherein the four WLEIS subscales in the self rater group were hypothesized to load on the first four latent factors, whereas the four WLEIS subscales in the peer rater group were hypothesized to load on the final four latent factors. Table 6 gives the correlations between the latent factors of both rater groups. Statistically significant correlations ranged between .16 and .35, which is in line with previous correlations between self-ratings and peer ratings (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988). The factor self emotion appraisal showed the weakest correlation between both rater groups

Table 4. Summary of Parameter Estimates for the Structural Noninvariance Model: Self Raters and Peer Raters

	Factor Pattern Coefficient				
Item	Self Emotion Appraisal	Others' Emotion Appraisal	Use of Emotion	Regulation of Emotion	Uniqueness
Has a good sense of why he/she has certain feelings most of the time	1.000	0	0	0	.202
Has good understanding of his/her own emotions	1.058	0	0	0	.095
Really understands what he/she feels	1.051	0	0	0	.165
Always knows whether or not he/she is happy	0.433	0	0	0	.479
Always knows his/her friends' emotions from their behavior	0	1.000	0	0	.216
Is a good observer of others' emotions	0	1.072	0	0	.137
Is sensitive to the feelings and emotions of others	0	0.620	0	0	.441
Has good understanding of the emotions of people around him/her	0	0.840	0	0	.242
Always sets goals for himself/herself and then tries his/her best to achieve them	0	0	1.000	0	.471
Always tells himself/herself he/she is a competent person	0	0	0.963	0	.608
Is a self-motivated person	0	0	1.604	0	.142
He/she would always encourage himself/herself to try his/her best	0	0	1.481	0	.169
Is able to control his/her temper and handle difficulties rationally	0	0	0	1.000	.272
Is quite capable of controlling his/her own emotions	0	0	0	1.016	.184
He/she can always calm down quickly when he/she is very angry	0	0	0	0.673	.593
Has good control of his/her own emotions	0	0	0	0.975	.205
	Factor Variances/Covariances				
Self Emotion Appraisal	.431*				
Others' Emotion Appraisal	.118*	.299* (.510*)			
Use of Emotion	.019	−.001	.137* (.216*)		
Regulation of Emotion	.092*	.113*	.068*	.474* (.629*)	

Note: Items were taken from Wong and Law (2002). Unstandardized solution in EQS is given. All factor pattern coefficients are statistically significant at $p < .05$. When parameters were not invariant, values for both groups are given. In that case, the values of the peer rater sample are within parentheses.
*$p < .05$.

Table 5. Tests of Measurement Invariance for the Multigroup Measurement Model of the Wong and Law Emotional Intelligence Scale Across Self Raters and Peer Raters (Replication Group)

Model	S-B χ^2	*df*	IFI	CFI	RMSEA	Model comparison	ΔS-B χ^2	Δ*df*	ΔCFI
1. Equal number of factors	236.945*	196	.973	.972	.040	—	—	—	—
2. Equal factor loadings	254.055*	208	.970	.969	.041	2 vs. 1	17.291	12	.003
3. Equal error variances	270.623*	224	.969	.968	.040	3 vs. 2	17.105	16	.001
4. Equal factor variances/covariances	296.945**	234	.958	.957	.045	4 vs. 3	26.368	10	.011

Note: S-B χ^2 = Satorra–Bentler chi-square; IFI = Bollen's incremental fit index; CFI = comparative fit index; RMSEA = root mean square error of approximation.
*$p < .05$. **$p < .01$.

Table 6. Factor Correlations Between the Self and the Peer Rating Groups ($N = 154$)

	Self Ratings			
Peer Ratings	1.	2.	3.	4.
1. Self Emotion Appraisal	.262***			
2. Others' Emotion Appraisal	.162*	.313***		
3. Use of Emotion	−.101	−.140	.348***	
4. Regulation of Emotion	−.048	−.088	−.018	.300***

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

($r = .26$, $p = .001$). However, differences between the latent factor correlations were not statistically significant.

Discussion

In recent years, the self-report approach to EI measurement has been criticized. Other reports might be used to side-step possible problems with self-reports of EI. However, no separate other report EI measures exist because existing self-report EI measures are also used in the context of other reports. This study tests the commonly made assumption that these extant EI measures are invariant across self-reports and other reports.

This study found evidence of configural invariance, which indicated that both rater groups use the same frame of reference when completing the WLEIS items. That is, they recognize these dimensions and agree that the WLEIS has four EI dimensions, each comprised of four items. So, we did not find a difference between the intrapersonal versus interpersonal dimensions of the WLEIS. In addition, there was evidence of metric invariance, implying that the self and other raters calibrate the WLEIS similarly (i.e., there are no differences in the scaling units). Both rater groups do not only agree on the four WLEIS dimensions and the items associated with these dimensions, they also agree on how the EI dimensions are manifested (Cheung, 1999). For example, both raters groups seem to consider an item such as "Always knows whether or not he/she is happy" the least appropriate one to indicate the Self Emotion Appraisal dimension because this item has the lowest factor loading associated with this EI dimension (see Table 5). Thus, self and peer raters concur on the importance of the items as indicators of their associated dimensions.

However, we did not find evidence for structural invariance as self-ratings of the WLEIS constructs appeared to have a more restricted range. This means that self raters use a narrower range of response intervals than other raters. A possible explanation for this might be that self raters want to present themselves according to social norms and standards (i.e., socially desirable responding; Zerbe & Paulhus, 1987). This tendency of individuals to present themselves favorably is in line with previous meta-analytic findings on 360-degree performance ratings that self-ratings are affected by egocentric biases (Conway & Huffcutt, 1997; Harris & Schaubroeck, 1988).

Our results were relatively robust as they were found across two different peer rater groups. At a practical level, these results indicate that knowledgeable others might serve as raters of extant EI measures. However, at the same time practitioners considering the use of the WLEIS in self and peer ratings should be aware that peer raters will use a wider range in the WLEIS dimensions and that self-ratings might be more biased than peer ratings. Thus, knowledgeable others might be used as a complement to self-ratings in EI measurement to obtain a more viable assessment.

One limitation of this study is that we investigated the measurement invariance across self and others of only one EI measure. Future research should examine whether our results generalize to other EI instruments. Similarly, other rater groups (colleagues, etc.) might be examined. As another interesting avenue for future research, research should examine the use of other reports in high-stakes selection contexts.

Authors' Note

A previous version of this paper was presented at the European Conference on Psychological Assessment (2009), Ghent.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Funding

This research was funded by a grant of the Fund for Scientific Research Flanders (grant number G.0348.07N [2701]).

References

Beaupré, M. G., Cheung, N., & Hess, U. (2000). *The Montreal Set of Facial Displays of Emotion* [Slides]. (Available from Ursula Hess, Department of Psychology, University of Quebec at Montreal, P.O. Box 8888, Station "Centre-ville," Montreal, Quebec H3C 3P8)

Bentler, P. M. (1995). *EQS: Structural equations program manual.* Encino, CA: Multivariate Software.

Boyatzis, R. E., Goleman, D., & Rhee, K. (2000). Clustering competence in emotional intelligence: Insights from the Emotional Competency Inventory (ECI). In R. Bar-On & J. Parker (Eds.), *Handbook of emotional intelligence* (pp. 343-362). San Francisco, CA: Jossey-Bass.

Brackett, M., Rivers, S., Shiffman, S., Lerner, N., & Salovey, P. (2006). Relating emotional abilities to social functioning: A comparison of self-report and performance measures of emotional intelligence. *Journal of Personality and Social Psychology, 91*, 780-795.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming.* Mahwah, NJ: Lawrence Erlbaum.

Charbonneau, D., & Nicol, A. (2002). Emotional intelligence and leadership in adolescents. *Personality and Individual Differences, 33*, 1101-1113.

Cheung, G. W. (1999). Multifaceted conceptions of self–other ratings disagreement. *Personnel Psychology, 52*, 1-36.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255.

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*, 331-360.

Davies, M., Stankov, L., & Roberts, R. D. (1998). Emotional intelligence: In search of an elusive construct. *Journal of Personality and Social Psychology, 75*, 989-1015.

Day, A. L., & Carroll, S. A. (2007). Faking emotional intelligence (EI): Comparing response distortion on ability and trait-based EI measures. *Journal of Organizational Behavior, 29*, 761-784.

Fan, X., & Thompson, B. (2001). Confidence intervals about score reliability coefficients, please: An EPM guidelines editorial. *Educational and Psychological Measurement, 61*, 517-531.

Freudenthaler, H. H., & Neubauer, A. C. (2005). Emotional intelligence: The convergent and discriminant validities of intra- and interpersonal emotional abilities. *Personality and Individual Differences, 39*, 569-579.

Grubb, W. L., & McDaniel, M. A. (2007). The fakability of Bar-On's Emotional Quotient Inventory Short Form: Catch me if you can. *Human Performance, 20*, 43-59.

Hancock, G. R. (1997). Structural equation modelling methods of hypothesis testing of latent variable means. *Measurement and Evaluation in Counseling and Development, 30*, 91-105.

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

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development, 34*, 177-189.

Jordan, P. J., Ashkanasy, N. M., Hartel, C. E., & Hooper, G. S. (2002). Workgroup emotional intelligence. Scale development and relationship to team process effectiveness and goal focus. *Human Resource Management Review, 12*, 195-214.

Law, K. S., Wong, C. S., Huang, G. H., & Li, X. (2008). The effects of emotional intelligence on job performance and life satisfaction for the search and development scientists in China. *Asia Pacific Journal of Management, 25*, 51-69.

Law, K. S., Wong, C. S., & Song, L. J. (2004). The construct and criterion validity of emotional intelligence and its potential utility for management studies. *Journal of Applied Psychology, 89*, 483-496.

Lopes, P. N., Grewal, D., Kadis, J., Gall, M., & Salovey, P. (2006). Evidence that emotional intelligence is related to job performance and affect and attitudes at work. *Psicothema, 18*, 132-138.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *The Mayer, Salovey, and Caruso Emotional Intelligence Test: Technical manual.* Toronto, Ontario, Canada: Multi-Health Systems.

Mayer, J. D., Salovey, P., Caruso, D. R., & Sitarenios, G. (2003). Measuring emotional intelligence with the MSCEIT V2.0. *Emotion, 3*, 97-105.

Paulhus, D. L., Lysy, D. C., & Yik, M. S. M. (1998). Self-report measures of intelligence: Are they useful as proxy IQ tests? *Journal of Personality, 66*, 525-553.

Pérez, J. C., Petrides, K. V., Furnham, A. (2005). Measuring trait emotional intelligence. In R. Schulze & R. D. Roberts (Eds.), *Emotional intelligence: An international handbook* (pp. 181-201). Cambridge, MA: Hogrefe & Huber.

Riordan, C. R., & Vandenberg, R. J. (1994). A central question in cross-cultural research: Do employees of different cultures interpret work-related measures in an equivalent manner? *Journal of Management, 20*, 643-671.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.

Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Dornheim, L. (1998). Development and validation of a measure of emotional intelligence. *Personality and Individual Differences, 25*, 167-177.

Shi, J., & Wang, L. (2007). Validation of emotional intelligence scale in Chinese university students. *Personality and Individual Differences, 43*, 377-387.

Song, L. J., Huang, G. H., Peng, K. Z., Law, K. S., Wong, C. S., & Chen, Z. J. (2010). The differential effects of general mental ability and emotional intelligence on academic performance and social interactions. *Intelligence, 38*, 137-143.

Van der Zee, K., Thijs, M., & Schakel, L. (2002). The relationship of emotional intelligence with academic intelligence and the Big Five. *European Journal of Personality, 16*, 103-125.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-69.

Wong, C. S., & Law, K. S. (2002). The effects of leader and follower emotional intelligence on performance and attitude: An exploratory study. *Leadership Quarterly, 13*, 243-274.

Zerbe, W. J., & Paulhus, D. L. (1987). Socially desirable responding in organizational-behavior: A reconception. *Academy of Management Review, 12*, 250-264.