Research Note

Applicant Versus Employee Scores on Self-Report Emotional Intelligence Measures

Filip Lievens,1* Ute-Christine Klehe,2* and Nele Libbrecht1

1Department of Personnel Management, Work and Organizational Psychology, Ghent University, Ghent, Belgium, 2Work and Organizational Psychology, University of Amsterdam, The Netherlands

Abstract. There exists growing interest to assess applicants’ emotional intelligence (EI) via self-report trait-based measures of EI as part of the selection process. However, some studies that experimentally manipulated applicant conditions have cautioned that in these conditions use of self-report measures for assessing EI might lead to considerably higher scores than current norm scores suggest. So far, no studies have scrutinized self-reported EI scores among a sample of actual job applicants. Therefore, this study compares the scores of actual applicants at a large ICT organization (n = 109) on a well-known self-report measure of EI to the scores of employees already working in the organization (n = 239). The current study is the first to show that applicants’ scores on a self-report measure of EI during the selection process are indeed higher (d = 1.12) and have less variance (SD ratio = 0.86/1) than incumbents’ scores. Finally, a meta-analytic combination of our results with those of earlier research showed that a score increase of about 1 SD in applicant conditions seems to be the rule, regardless of the type of setting, self-report EI measure, and within-versus between-subjects design employed.

Keywords: score inflation, emotional intelligence, personnel selection, trait-based emotional intelligence self-report, applicants

Emotional intelligence (EI) is a topic of high interest and of debate among both practitioners and academicians in terms of its ability to predict employee well-being (Schutte, Malouf, Thorsteinsson, Bhullar, & Rooke, 2007) and performance in both team (Bell, 2007) and individual contexts (Joseph & Newman, 2010). As practitioners seem eager to assess the EI of potential new employees in order to make selection decisions (Zeidner, Roberts, & Matthews, 2004), various test vendors have started to sell EI measures (Joseph & Newman, 2010). Although some of these measures are ability tests (Mayer, Salovey, & Caruso, 2002), most popular measures in practice are self-report trait-based EI measures wherein respondents report their “abilities, competencies, and skills related to understanding oneself and others, relating to peers and family members, and adapting to changing environmental situations and demands” (Bar-On, 2002).

Despite the growing interest to use self-report EI measures in selection practice, they might suffer from two problems. The first one is a poor definition of the construct. Many self-report EI measures show considerable conceptual and empirical overlap with traditional personality measures (Davies, Stankov, & Roberts, 1998; Dawda & Hart, 2000; Newsome, Day, & Catano, 2000; Schutte et al., 1998). Second, unlike ability-based measures (Day & Carroll, 2008), use of self-report measures for assessing EI among applicants might lead to considerably higher scores than current norm scores typically indicate in the manual of the EI measures.

Several studies have aimed to examine this second issue (Christiansen, Janovics, & Siers, 2010; Day & Carroll, 2008; Grubb & McDaniel, 2007; Whitman, Van Rooy, Viswesvaran, & Alonso, 2008). Most of these studies were laboratory studies with two experimental conditions: One wherein students were asked to respond honestly and one wherein they were asked to respond as favorably as possible or as if they were applying for a job. Another study (Engelberg & Sjöberg, 2005) compared responses of business school candidates to responses given by an anonymous group of current students. All these prior studies found that under experimentally manipulated “applicant” conditions, participants ascribed themselves an EI about 0.5 to 1 standard deviations above the EI reported under standard nonevaluative or “honest” conditions, which might affect applicants’ likelihood of being selected on the basis of EI scores (see Day & Carroll, 2008). Whitman et al. (2008) further showed that this effect was considerably larger in within-subjects comparisons than it was in between-subjects comparisons and both Day and Carroll (2008) and Engelberg and Sjöberg

* The order of authorship was decided on the basis of a completely random yet skillfully executed flip of a perfectly normal coin.

© 2011 Hogrefe Publishing

Journal of Personnel Psychology 2011; Vol. 10(2):89–95
DOI: 10.1027/1866-5888/a000036
related this increase in EI scores to impression management. Finally, the correlation matrixes in both Day and Carroll (2008) and Grubb and McDaniel (2007) suggested that applicant-like conditions lead to inflated inter-scale correlations.

In sum, these prior studies suggest that trait-based EI measures can lead to higher scores under experimentally-induced applicant conditions. What needs further study, however, is how much real job applicants do report higher EI scores. Otherwise, relying on regular EI norm-data might lead to a gross overestimation of the EI actually present in an applicant sample. In terms of triangulating research methods (Mesmer-Magnus & Viswesvaran, 2006), this study is the first to examine real job-applicants’ scores on a well-known self-report EI measure. As a benchmark, we compare the EI scores of those applicants to scores of current employees of the same organization. We take EI scores of employees as a benchmark because it can be assumed that employees have no situational demands to make a positive impression, although at the same time we acknowledge that alternative explanations such as differences in job experience and organizational level may limit the interpretation of applicant versus employee differences as mere response distortion (Mesmer-Magnus & Viswesvaran, 2006).

That such triangulation is necessary is evidenced by research about score increases on personality inventories, which are conceptually and empirically related to self-report trait-based EI measures. Numerous laboratory studies showed that – when experimentally induced to do so – people report higher scores on self-report personality inventories (Viswesvaran & Ones, 1999). In addition, there was less variance in their scores. Finally, results often revealed a decrement in the personality measure’s construct-related validity due to higher interrelationships among personality traits (Schmit & Ryan, 1993). Other personality research, however, has called for caution, pointing out that most of the aforementioned results are based on laboratory studies with instructed applicant or even “faking” conditions that create a worst case scenario. Among real applicants, however, scores on personality inventories are often just somewhat or not at all higher than among job incumbents (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Ellingson, Sackett, & Connelly, 2007). Indeed, the finding that people can report higher scores when instructed to do so does not imply that they will report higher scores when applying for a job (McFarland & Ryan, 2006). We still do not know whether this conclusion generalizes to job applicants completing trait-based self-report EI measures.

This study’s objective was to compare real job-applicants’ and job-incumbents’ responses on a self-report trait-based EI measure. Extrapolating from the personality literature we hypothesize the following:

Hypothesis 1: Applicants will report higher levels of EI than incumbents on a self-reported EI measure.

Hypothesis 2: Applicants’ self-reported EI scores will show a reduced variance as compared to the EI reported by incumbents.

Hypothesis 3: As compared to incumbents’ EI scores, applicants’ self-reported EI scores will show higher covariances among the different components of EI.

Method

Sample

Data were collected in two samples. The applicant sample consisted of 109 job applicants (52% male) who applied for a job in a large ICT organization. Applicants’ ages ranged from 18 to 60 years (M = 29, SD = 8) and 46% had a higher education degree. The incumbent sample consisted of 239 volunteer incumbents in the same organization (43% male). Incumbents’ ages ranged from 22 to 63 years (M = 43, SD = 9) and 41% had a university degree. We tried to match our samples as much as possible by collecting data for incumbents and applicants in the same departments, excluding higher management incumbents from the study, as no higher functions were being sought for in the applicant sample.

Procedure

The data collection was approved by the HR department of the company. As part of the hiring process, the applicants completed the organization’s selection test battery including the EI measure. Only upon completion of the battery were applicants informed that the EI measure would not be used in making selection decisions. After this briefing, all applicants were asked to sign an informed consent.

Incumbents were assured that their responses would be used only for research purposes by the researchers and were asked to complete the EI measure at work during office hours. An e-mail containing a link to the web-based EI administration was sent to 427 incumbents. Two weeks later a reminder was e-mailed. Study participation was voluntary. The response rate was about 56%.

Wong and Law Emotional Intelligence Scale

EI was measured via the Wong and Law (2002) Emotional Intelligence Scale (WLEIS), which is a four-dimensional self-report EI measure designed to assess EI in the workplace. Four items each assess the four EI dimensions identified by Davies, Stankov, and Roberts (1998): The Self-Emotion Appraisal dimension (e.g., “I really understand what I feel”) assesses individuals’ ability to understand and express their own emotions. The Other’s Emotion Appraisal dimension (e.g., “I always know my friends’ emotions from their behavior”) measures peoples’ ability to perceive and understand the emotions of others. The Use of Emotion dimension (e.g., “I always tell myself...
I am a competent person”) denotes individuals’ ability to use their emotions effectively by directing them toward constructive activities and personal performance. Finally, the Regulation of Emotion dimension (e.g., “I have good control of my own emotions”) refers to individuals’ ability to regulate their own emotions. The WLEIS was measured with a 5-point Likert scale, ranging from 1 (totally disagree) to 5 (totally agree).

Previous research has found support for the underlying four-factor structure, reliability, convergent, and discriminant validity of the WLEIS (Law, Wong, Huang, & Li, 2008; Law, Wong, & Song, 2004; Shi & Wang, 2007; Wong & Law, 2002). WLEIS scores have also shown validity for predicting life satisfaction, academic performance, job performance, and job satisfaction (Law et al., 2008; Song et al., 2010; Sy, Tram, & O’Hara, 2006; Wong & Law, 2002). In summary, the WLEIS represents one of the most stringently developed and validated self-report measures on EI available to date.

Results

Table 1 presents the means, standard deviations, and intercorrelations of the WLEIS subscales among incumbents (top) and applicants (bottom). Hypothesis 1 proposed that on average applicants will ascribe themselves higher levels of EI than will incumbents. As shown in Table 2, this was true for all four dimensions of the WLEIS. The larger effect sizes for the dimensions use of emotions and regulation of emotions might be due to the high conceptual overlap of these EI dimensions with the personality dimensions conscientiousness and emotional stability, which are typically the personality dimensions most inflated among applicants (Birkeland et al., 2006). Overall, applicants reported an EI that was 1.12 SD higher than that reported by incumbents (see also Figure 1). Additional analyses showed that these differences between applicants and incumbents could not be explained by differences in age and gender between these two groups (Table 3).

While these results seem to align with the earlier studies on this phenomenon (Christiansen et al., 2010; Day & Carroll, 2008; Engelberg & Sjöberg, 2005; Grubb & McDaniel, 2007; Whitman et al., 2008) conducted in laboratory or educational settings, we additionally ran a small random effects meta-analysis across these studies and our own dataset, using the meta-analytic program provided in Hunter and Schmidt (1990). A bare bones meta-analysis (k = 7), weighting effects by sample size (total n of comparisons = 807), revealed an average observed effect size of $d = 0.88$ with SD = 0.12 (controlling for the unreliability of the measures.

Table 1. Means, standard deviations, internal consistency reliabilities, and correlations between the scales in the incumbent and the applicant samples

<table>
<thead>
<tr>
<th>Dimension</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>M</th>
<th>SD</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>−.10</td>
<td>.10</td>
<td>.01</td>
<td>−.03</td>
<td>.00</td>
<td>.03</td>
<td>42.72</td>
<td>9.12</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>−.22*</td>
<td>.09</td>
<td>.11</td>
<td>.08</td>
<td>−.14</td>
<td>.03</td>
<td>0.57</td>
<td>0.50</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Self-emotion appraisal</td>
<td>−.04</td>
<td>.02</td>
<td>.46**</td>
<td>.25**</td>
<td>.24**</td>
<td>.71**</td>
<td>3.97</td>
<td>0.61</td>
<td>.83</td>
<td></td>
</tr>
<tr>
<td>Others’ emotion appraisal</td>
<td>−.05</td>
<td>.15</td>
<td>.21**</td>
<td>.15**</td>
<td>.15**</td>
<td>.63**</td>
<td>3.67</td>
<td>0.58</td>
<td>.77</td>
<td></td>
</tr>
<tr>
<td>Use of emotion</td>
<td>−.04</td>
<td>.08</td>
<td>.30**</td>
<td>.15**</td>
<td>.15**</td>
<td>.63**</td>
<td>3.67</td>
<td>0.58</td>
<td>.77</td>
<td></td>
</tr>
<tr>
<td>Regulation of emotion</td>
<td>−.19*</td>
<td>.01</td>
<td>.41**</td>
<td>.31**</td>
<td>.30**</td>
<td>.69**</td>
<td>3.29</td>
<td>0.77</td>
<td>.83</td>
<td></td>
</tr>
<tr>
<td>WLEIS total</td>
<td>−.03</td>
<td>.10</td>
<td>.70**</td>
<td>.58**</td>
<td>.68**</td>
<td>.74**</td>
<td>3.63</td>
<td>0.42</td>
<td>.81</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>28.52</td>
<td>.48</td>
<td>4.19</td>
<td>3.91</td>
<td>4.15</td>
<td>4.09</td>
<td>4.08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>8.37</td>
<td>.50</td>
<td>.51</td>
<td>.48</td>
<td>.60</td>
<td>.53</td>
<td>.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>–</td>
<td>–</td>
<td>.74</td>
<td>.57</td>
<td>.69</td>
<td>.68</td>
<td>.79</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Top triangle: Employees (n = 239). Lower triangle: Applicants (n = 109). *p < .05. **p < .01. Gender: 0 = men; 1 = women.

Table 2. Means comparisons between incumbents and applicants

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Employees (n = 239)</th>
<th>Applicants (n = 109)</th>
<th>SD ratio (applicants/employees)</th>
<th>t (df = 346)</th>
<th>d</th>
<th>95% confidence interval around d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-emotion appraisal</td>
<td>3.97</td>
<td>0.61</td>
<td>4.19</td>
<td>0.51</td>
<td>0.84</td>
<td>3.36**</td>
</tr>
<tr>
<td>Others’ emotion appraisal</td>
<td>3.67</td>
<td>0.58</td>
<td>3.91</td>
<td>0.48</td>
<td>0.83</td>
<td>4.01**</td>
</tr>
<tr>
<td>Use of emotion</td>
<td>3.58</td>
<td>0.57</td>
<td>4.15</td>
<td>0.60</td>
<td>1.05</td>
<td>8.46**</td>
</tr>
<tr>
<td>Regulation of emotion</td>
<td>3.29</td>
<td>0.77</td>
<td>4.09</td>
<td>0.53</td>
<td>0.69</td>
<td>11.16**</td>
</tr>
<tr>
<td>WLEIS total</td>
<td>3.63</td>
<td>0.42</td>
<td>4.08</td>
<td>0.36</td>
<td>0.86</td>
<td>9.75**</td>
</tr>
</tbody>
</table>

Note. **p < .01.
the latent effect would be \( d = 0.94, SD = 0.15 \). The 80% confidence interval of the data ranged from .72 to 1.03, giving further credibility to the strength of the effect across studies. The proportion of the observed variance in \( d \) (0.053) due to artifacts (0.038) was 73%, thus suggesting a considerable generalizability of results across these lab and field studies relying on within- as well as between-person designs. A set of separate analyses for within and between designs both revealed the same average \( d \) of 0.87 (within designs, \( k = 3 \))/0.88 (between designs, \( k = 4 \)), lending further support for Hypothesis 1.

Hypothesis 2 proposed that applicants’ self-reported EI scores will have reduced variance compared to those of incumbents. Figure 1. Score increases among applicants (stripped line) and incumbents (solid line) on the global EI scale (Panel 1, depicting percentage of respondents by average score on the 5-point scale) and EI subscales (Panel 2, depicting average score on the 5-point scale by the WLEIS dimensions: SEA: Self-emotion appraisal; OEA: Others’ emotion appraisal; UOE: Use of emotion; ROE: Regulation of emotion).
incumbents. We tested this assumption via multiple group comparisons in structural equation modeling, using AMOS 16 (Arbuckle, 2003). An unconstrained model as well as a measurement invariant model (i.e., constraining factor loadings to be equal across incumbents and applicants) fitted the data reasonably well (Table 4). As soon as we constrained the factor variances to be equal, however, the model showed a significant decrement in fit. A critical ratio test (Table 5) revealed significantly smaller factor variances among applicants on all but one of the four dimensions tested, thereby largely supporting Hypothesis 2. Thus, applicants’ responses resulted in less variance from which future outcomes might be predicted.

Finally, Hypothesis 3 proposed that applicants’ self-reported EI scores will show inflated covariances among the EI dimensions, compared to incumbents’ self-reported EI scores. Neither the correlational results (Table 1) nor a model-comparison constraining factor covariances to be equal across incumbents and applicants (Table 3) supported this hypothesis.

Discussion

Past personality research has suggested that laboratory findings might overestimate the degree to which actual job applicants report higher scores than incumbents. In particular, field research has indicated that applicants score higher on only some instead of all personality dimensions (Birkeland et al., 2006). This study’s results demonstrate that this overestimation does not generalize to self-reported EI scores. Our current findings of score increases on a well-established EI measure closely mirror those found in prior studies conducted with a variety of self-report EI measures in both within- and between-person designs in laboratory and educational settings (Christiansen et al., 2010; Day & Carroll, 2008; Engelberg & Sjöberg, 2005; Grubb & McDaniel, 2007; Whitman et al., 2008). Triangulation and a meta-analytic combination of our results with those of earlier lab and field studies show a consistent picture: applicant self-reported EI scores are about a standard deviation above those reported by comparable incumbents in the organization. While clearly based on small k's, score increases in applicant conditions thus seem to be the rule, regardless of the type of setting (laboratory or field setting), self-report EI measure used (e.g., the EQ-i, EQ-i: S, SREIT, TMMS, WLEIS), and within- versus between-subjects design employed.

Apart from the fact that applicants can and do ascribe themselves considerably higher EI on self-report measures in a personnel selection process, there was also evidence of reduced variance. Hence, a primary follow-up question for future research is whether the variance reduction of applicants’ self-reported EI will also impair the criterion-related validity of the EI measure when used for predicting performance (Bell, 2007; Joseph & Newman, 2010).

This study is not without limitations. Unlike within-person designs, which have already been used in different laboratory studies with comparable results (Day & Carroll, 2008; Grubb & McDaniel, 2007; Whitman et al., 2008), a

---

Table 4. Summary of model fit results of measurement models

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>χ²</th>
<th>p</th>
<th>χ²/df</th>
<th>CFI</th>
<th>TLI</th>
<th>IFI</th>
<th>RMSEA</th>
<th>90% CI</th>
<th>Model comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incumbents</td>
<td>98</td>
<td>160.98</td>
<td>.00</td>
<td>1.64</td>
<td>.95</td>
<td>.94</td>
<td>.95</td>
<td>.05</td>
<td>.04–.07</td>
<td>Model 2 – 1 12 8.68</td>
</tr>
<tr>
<td>Applicants</td>
<td>98</td>
<td>127.44</td>
<td>.02</td>
<td>1.30</td>
<td>.95</td>
<td>.93</td>
<td>.94</td>
<td>.05</td>
<td>.02–.08</td>
<td>Model 3 – 1 16 32.70</td>
</tr>
<tr>
<td>1. Unconstrained model</td>
<td>196</td>
<td>288.61</td>
<td>.00</td>
<td>1.47</td>
<td>.95</td>
<td>.94</td>
<td>.95</td>
<td>.04</td>
<td>.03–.05</td>
<td>Model 4 – 1 18 18.77</td>
</tr>
<tr>
<td>2. Equal factor loadings</td>
<td>208</td>
<td>297.29</td>
<td>.00</td>
<td>1.45</td>
<td>.95</td>
<td>.94</td>
<td>.95</td>
<td>.04</td>
<td>.03–.04</td>
<td>Model 5 – 2 6 10.09</td>
</tr>
<tr>
<td>3. Equal factor variances</td>
<td>212</td>
<td>321.31</td>
<td>.00</td>
<td>1.52</td>
<td>.94</td>
<td>.93</td>
<td>.94</td>
<td>.04</td>
<td>.03–.04</td>
<td>Model 6 – 2 4 24.02</td>
</tr>
<tr>
<td>4. Equal factor covariances</td>
<td>214</td>
<td>307.38</td>
<td>.00</td>
<td>1.44</td>
<td>.95</td>
<td>.93</td>
<td>.94</td>
<td>.04</td>
<td>.03–.04</td>
<td>Model 7 – 2 6 12</td>
</tr>
</tbody>
</table>

Note. CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; IFI = Incremental Fit Index; RMSEA = root mean square error of approximation.
between-subjects design as used in the current study always leaves the door open to alternative possible explanations for the differences found between the applicant and the incumbent samples (though see Mesmer-Magnus & Viswesvaran, 2006 for similar concerns regarding within-subject field-study designs). In the current study, we ruled out the effects of age and gender. Also, while we did not measure job experience or organizational level as possible covariates, we do not believe that either might have accounted for the differences found: The older incumbent sample likely had considerably more experience and held a comparable or higher organizational level than the applicant sample. Experience and organizational level, in turn, correlate positively with different ability and personality facets (e.g., Ones & Dilchert, 2009) which are conceptually and empirically positively related to trait-based EI (Joseph & Newman, 2010). Despite this, however, the older incumbent sample still reported significantly lower EI. Additionally, while external job candidates tend to ascribe themselves somewhat more favorable personality score than internal job candidates (Ones & Viswesvaran, 2007), such effects are far too small to account for the considerably higher self-reported EI among applicants than incumbents. That said, we encourage further research to consider the effects of general variables such as personality and EI and more specific variables such as job experience, source of applicants, and occupational level on score increases on self-report EI measures.

At a practical level, our results provide a warning signal to practitioners who want to implement self-report EI measures in their selection practice. This study shows that in that case norms on self-reported EI (gathered in research contexts) in the manual will not generalize to applicant samples. In addition, caution should be exerted when selecting applicants on the basis of their self-reported EI. At the very least, applicant norms on self-report EI measures are needed for appropriate use of such measures in employee selection.

References


Ute-Christine Klehe

Arbeids- & Organisatiepsychologie
Universiteit van Amsterdam
Roetersstraat 15
1018 WB Amsterdam
The Netherlands
Tel. + 31 20 525-6745
Fax + 31 20 639-0531
E-mail u.klehe@uva.nl