

Singapore Management University Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of
Business

Lee Kong Chian School of Business

12-2012

Validity evidence for the situational judgment test paradigm in emotional intelligence measurement

Nele LIBBRECHT
Ghent University

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg
DOI: <https://doi.org/10.1080/00207594.2012.682063>

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

Part of the [Industrial and Organizational Psychology Commons](#), and the [Organizational Behavior and Theory Commons](#)

Citation

LIBBRECHT, Nele and LIEVENS, Filip. Validity evidence for the situational judgment test paradigm in emotional intelligence measurement. (2012). *International Journal of Psychology*. 47, (6), 438-447. Research Collection Lee Kong Chian School Of Business. **Available at:** https://ink.library.smu.edu.sg/lkcsb_research/5612

This Journal Article is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Validity evidence for the situational judgment test paradigm in emotional intelligence measurement

Nele Libbrecht and Filip Lievens

Department of Personnel Management, Work and Organizational Psychology, Ghent University, Ghent, Belgium

To date, various measurement approaches have been proposed to assess emotional intelligence (EI). Recently, two new EI tests have been developed based on the situational judgment test (SJT) paradigm: the Situational Test of Emotional Understanding (STEU) and the Situational Test of Emotion Management (STEM). Initial attempts have been made to examine the construct-related validity of these new tests; we extend these findings by placing the tests in a broad nomological network. To this end, 850 undergraduate students completed a personality inventory, a cognitive ability test, a self-report EI test, a performance-based EI measure, the STEU, and the STEM. The SJT-based EI tests were not strongly correlated with personality and fluid cognitive ability. Regarding their relation with existing EI measures, the tests did not capture the same construct as self-report EI measures, but corresponded rather to performance-based EI measures. Overall, these results lend support for the SJT paradigm for measuring EI as an ability.

Keywords: Emotional intelligence; Situational judgment tests; Emotional intelligence assessment.

Jusqu'à ce jour, plusieurs approches ont été proposées pour évaluer l'IE. Récemment, deux nouveaux tests d'IE basés sur le paradigme du test de jugement situationnel (TJS) ont été produits, soit le *Situational Test of Emotional Understanding* (STEU) et le *Situational Test of Emotion Management* (STEM). Nous ajoutons aux résultats des premières tentatives qui ont été menées pour étudier la validité de construit de ces nouveaux tests en les situant dans un large réseau nomologique. À cette fin, 850 étudiants du premier cycle ont répondu à un inventaire de personnalité, à un test d'aptitude cognitive, à un test auto-rapporté d'IE, à une mesure d'IE basée sur la performance, au STEU et au STEM. Les tests d'IE basés sur le TJS n'ont pas donné lieu à une corrélation importante avec la personnalité et l'aptitude cognitive fluide. En ce qui concerne leur relation avec les mesures existantes de l'IE, les tests n'ont pas mis en évidence le même construit que les mesures auto-rapportées d'IE, mais se rapprochaient plutôt des mesures d'IE basées sur la performance. Dans l'ensemble, ces résultats confortent l'utilisation du paradigme du TJS pour la mesure de l'IE en tant qu'aptitude.

Hasta la fecha se ha propuesto diversas aproximaciones de medición para evaluar la EI. Recientemente se ha desarrollado dos nuevas pruebas de EI basadas en el paradigma de la prueba de juicio situacional (SJT), la Prueba Situacional de Entendimiento Emocional (STEU) y la Prueba Situacional de Manejo de Emociones (STEM). A pesar de que se ha realizado intentos iniciales para examinar la validez relacionada con el constructo de estas nuevas pruebas, extendimos estos hallazgos al situar estas pruebas en una amplia red nomológica. Para este fin, 850 estudiantes de pregrado completaron un inventario de personalidad, una prueba de habilidad cognoscitiva, un autoinforme de EI, una medida de EI basada en la ejecución, la STEU y la STEM. Las pruebas de EI basadas en la STJ no estuvieron fuertemente correlacionadas con la personalidad ni con la habilidad

In recent years, the concept of emotional intelligence (EI; Mayer & Salovey, 1997; Salovey & Mayer, 1990) has received considerable attention in both the academic and the practitioner literature (Joseph & Newman, 2010; Mayer, Roberts, & Barsade, 2008) and has been proposed as the newest alternative for predicting a wide range of constructs over and above cognitive ability and achievement in employment and educational contexts. However, despite the research interest and the potential potency of EI as an individual difference predictor, there are still many disagreements about the definition and measurement of EI.

This study focuses on the measurement of EI and concentrates on the use of situational judgment tests (SJTs) as a recent paradigm (MacCann & Roberts, 2008; Wilhelm, 2005). SJTs measure an individual's judgment concerning situations by presenting examinees with scenarios and asking them to identify an appropriate response (McDaniel, Hartman, Whetzel, & Grubb, 2007; Motowidlo, Dunnette, & Carter, 1990; Weekley, Ployhart, & Holtz, 2006). In this study, we extend construct-related validity evidence of the SJT measurement approach by placing it in a network with other EI, cognitive ability, and personality measures.

EMOTIONAL INTELLIGENCE MEASUREMENT PARADIGMS

A number of measurement approaches exist for assessing EI. A first approach is the use of self-reports wherein individuals report their own level of EI and typically assess EI as defined within the trait EI model as "a constellation of emotional self-perceptions located at the lower levels of personality hierarchies" (Petrides, 2010, p. 137). Second, other-reports have been used to measure EI (Law, Wong, & Song, 2004) wherein knowledgeable informants indicate how well the scale items describe the focal person's ability. A third approach is the use of performance-based tests wherein individuals indicate the appropriate reaction to emotion-related problems or identify facial expressions. The answers are evaluated against predetermined scoring criteria (Freudenthaler & Neubauer, 2005). These tests are consistent with a definition of "ability EI," namely EI as a broad

intellective factor comprising four conceptually related hierarchical branches of emotion-related abilities (i.e., perceiving, using, understanding, and managing emotions; Mayer & Salovey, 1997). The most widely used ability EI measure is the Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT; Mayer, Salovey, & Caruso, 2002). Other tests (e.g., the Montreal Set of Facial Displays of Emotion; Beaupré, Cheung, & Hess, 2000) measure the specific EI ability to perceive emotions.

So far, there has been a paucity of alternatives to assess other EI abilities, such as the abilities to understand emotions correctly and manage emotions effectively. New approaches are sorely needed to assess emotional abilities because "a diversity of methods ensures that researchers can accurately interpret research findings as method-related or construct-related" (MacCann & Roberts, 2008, p. 550). Developing new EI measures also broadens the measurement of EI, thereby getting away from the sole reliance on either self-report or performance tests. Recently, the SJT paradigm has been proposed as such a complementary approach for measuring EI.

ASSESSING EMOTIONAL INTELLIGENCE WITH SITUATIONAL JUDGMENT TESTS

The SJT approach is a popular measurement approach in personnel selection (Weekley & Ployhart, 2006). Christian, Edwards, and Bradley's recent meta-analysis (2010) revealed that SJTs are most often used to assess leadership and interpersonal skills. As it is also possible to insert emotion-related scenarios into SJT items, Schulze, Wilhelm, and Kyllonen (2007) posited that "the prospects for SJTs are very promising for an improved assessment of EI" (p. 221). So far, initial steps have been taken to use the SJT format for measuring EI. For instance, some subtests of the MSCEIT (e.g., Emotion Management), the Levels of Emotional Awareness Scale (Lane, Quinlan, Schwartz, Walker, & Zeitlin, 1990), and the Test of Emotional Intelligence (Blickle et al., 2009) consist of SJT-like items. MacCann and Roberts (2008) have created two full-blown paper-and pencil EI SJTs. First, the Situational Test of Emotional

Understanding (STEU) was designed to measure individuals' ability to identify the emotions that most likely result from specific situations. The development of the STEU (item construction, response option generation, and scoring) was based on Roseman's appraisal theory (2001). To create and objectively score the STEU, MacCann and Roberts (2008) developed situations for 14 emotions to represent specific combinations of the appraisal dimensions. Second, the Situational Test of Emotion Management (STEM) was developed to measure individuals' ability to effectively manage specific emotional situations by using the correct strategies as identified by an expert group (i.e., researchers, psychologists, and counselors).

A key outstanding issue is how the STEU and the STEM relate to cognitive ability and personality. In light of their content-related description (emphasis on people's *understanding* which emotions occur in STEU and applying emotion *knowledge* in STEM), their use of knowledge response instructions (e.g., "what is the best answer?" instead of "what are you most likely to do?" as in behavioral tendency response instructions), and a multiple choice response format, both tests were developed to measure EI as a set of mental abilities rather than preferred ways of behaving. Hence, they should relate to pre-existing intelligence measures and be relatively independent from personality trait inventories wherein people report on their behavioral tendencies (Mayer, Caruso, & Salovey, 1999).

To our knowledge, only two studies have examined the validity of these new EI SJTs (Austin, 2010; MacCann & Roberts, 2008). To examine whether the SJTs assess cognitive processing of affective information and can be considered part of the intelligence rather than personality domain, MacCann and Roberts (2008) scrutinized the correlations between the STEU and the STEM scores and personality. With the highest correlation at .24, their results demonstrated that the EI tests were distinct from personality. Next, they found moderate relationships for the STEU and the STEM with a vocabulary test of verbal intelligence ($r = .49$ and $r = .41$, respectively), indicating that the tests are related to some aspects of cognitive ability. However, vocabulary tests are generally considered to measure crystallized intelligence (Gc; acquired acculturated knowledge) rather than fluid intelligence (Gf; innate reasoning ability) (Cattell, 1987), and, thus, it remains unknown how the SJTs would correlate with Gf measures. Recently, Austin (2010) examined the associations of the STEU and the STEM with a letter series test as Gf measure and found

no relations. Regarding the relations of the STEU and the STEM with Gc, Austin replicated the relation between the STEU and the vocabulary test used by MacCann and Roberts (2008) ($r = .32$) but not the relation between the STEM and vocabulary scores.

Another critical question concerns how the SJTs correlate with existing EI measures. It is important to rule out the possibility that EI SJTs capture people's self-judged emotional abilities. As both EI SJTs are tests that assess whether respondents *know* important information, they should not correspond highly to self-report EI scales in which people evaluate their own abilities, because self-assessments of skills are often flawed (Dunning, Heath, & Suls, 2004). To demonstrate that the SJTs assess abilities, they should be correlated with performance-based EI tests. In a first attempt to answer this question, Austin (2010) examined the associations of the SJT EI tests with existing performance-based and self-report trait EI measures. Her results showed positive correlations between the STEU and the STEM and MSCEIT total scores ($r = .33$ and $r = .36$, respectively), positive correlations between the STEU and emotion perception tasks (average $r = .26$), and no significant relations for the STEU and the STEM with trait EI tests (average $r = .00$ and $r = .13$, respectively).

PRESENT STUDY AND HYPOTHESES

This study aims to extend the scarce construct-related validity evidence for the STEU and the STEM. If SJTs are a feasible complementary measurement paradigm for EI assessment, an important benefit will be that EI measurement is broadened. However, before researchers can compare self-report, other-report, performance-based, and SJT paradigms to expand our knowledge of the assessed construct EI (by disentangling the method and construct effects), more insight in the nomological network of the different measurement paradigms is needed. Therefore, this study aims to expand the information available on the STEU and the STEM by placing them in a broad nomological network, including a personality inventory, a broad Gf measure, a self-report EI test based on the well-established four-branch ability EI model, and a performance-based EI measure assessing the emotion perception ability. On the basis of the abovementioned theoretical notions behind the development of the EI SJTs and prior validation research (Austin,

2010; MacCann & Roberts, 2008) we posit the following.

- *H1*: There will be small correlations between the STEU and the STEM on one hand and the Big Five personality traits on the other.
- *H2*: There will be moderate correlations between the STEU and the STEM on one hand and Gf on the other.
- *H3*: The STEU and the STEM will correlate more strongly with a performance-based EI test than with a self-report EI measure.

METHOD

Participants and procedure

Eight hundred and fifty students (37% male, mean age = 20.2 years), who were in their first (25%), second (26%), third (27%), or fourth (21%) year of medical school, participated for partial completion of a course requirement. Prior to entering medical college, all students had passed the Flemish medical admission exams. In PC-equipped rooms, participants completed the computer-based test battery containing demographic, personality, self-report EI, performance-based EI, and the two SJT EI measures. About 10% of the sample ($N=86$) did not complete the full test battery. Results of statistical comparisons between the available responses of the partially responding students and those of respondents who completed the whole survey (Wagner & Kemmerling, 2010) showed that nonresponse bias was unlikely to alter our results.

Measures

Situational Test of Emotional Understanding (STEU)

Participants completed the 42 multiple-choice items of the STEU (MacCann & Roberts, 2008): 14 items were context-reduced, 14 had a workplace content, and 14 had a personal-life context. The original STEU was translated into Dutch/Flemish by one of the authors and an English–Dutch translator. This translated version was checked by an emotion expert. No modifications needed to be made. Respondents had to identify which emotion would most likely result from each described situation. Correct answers were determined using Roseman's (2001) appraisal theory. A sample item is (correct answer in parentheses): *There is great weather on the day Jill is going on an outdoor picnic.*

Jill is most likely to feel? (a) pride [(b) joy] (c) relief (d) guilt (e) hope. Given the multidimensional nature of SJTs, internal consistency is not a good reliability measure for them and test–retest reliability has been suggested as a better measure (Whetzel & McDaniel, 2009). Test–retest reliability was assessed in a pilot study where the STEU was administered on a two-week interval among 32 psychology students (75% women) and was .72.

Situational Test of Emotion Management (STEM)

The 30-item version of the STEM (MacCann & Roberts, 2008) was administered in multiple-choice format and was scored according to expert mean ratings. The STEM was translated into Dutch/Flemish via the same procedure as the STEU. In each item, an emotional situation was presented and participants had to select the most effective reaction to manage both the emotions and problems faced in that situation. A sample item is (answer with highest score in parentheses): *Alan helps Trudy, a peer he works with occasionally, with a difficult task. Trudy complains that Alan's work isn't very good, and Alan responds that Trudy should be grateful he is doing her a favor. They argue. What action would be the most effective for Alan? (a) Apologize to Trudy, (b) Stop helping Trudy and don't help her again, (c) Try harder to help appropriately, [(d) Diffuse the argument by asking for advice].* The STEM was also administered in our abovementioned pilot study and test–retest reliability was .85.

Self-rated emotional intelligence

To measure EI with a self-report measure, we used a short form of the Self-Rated Emotional Intelligence Scale (SREIS) developed by Brackett, Rivers, Shiffman, Lerner, and Salovey (2006) because it measures all four ability EI branches. The SREIS contains four items for perceiving emotions ($\alpha = .67$), three items for using emotions ($\alpha = .84$), four items for understanding emotions ($\alpha = .84$), and four items for managing emotions of others ($\alpha = .72$). Participants used a five-point Likert scale, ranging from 1 (*very inaccurate*) to 5 (*very accurate*), to indicate the extent to which each statement described them.

Emotion perception

To measure EI via a performance-based test, we selected 48 morphed facial expressions of anger, joy, sadness, fear, disgust, and shame

from the Montreal Set of Facial Displays of Emotion (MSFDE, Beaupré et al., 2000) because this emotion perception task assesses individuals' ability to identify facial emotional expressions, which is a well-established ability within the EI domain. Each photograph was presented on a computer screen for 1 s, followed by a screen on which participants were asked to select which emotion was displayed. On the base of six repeated items we calculated the intrarater reliability of this scale's score, which was .68.

Personality

Each of the Big Five Personality traits was assessed with 9 of the 10 Likert-type items (1 = *very inaccurate* and 5 = *very accurate*) of the International Personality Item Pool (Goldberg, 1999). Nine instead of ten items were measured because in our confirmatory factor analysis (CFA) analyses we used an item-parceling procedure (Little, Cunningham, Shahar, & Widaman, 2002) that randomly divided the nine items for each personality trait into three subscale composites. The item-parceling procedure was used because the number of personality factors is small relative to the number of observed variables (Drasgow & Kanfer, 1985). Internal consistency reliabilities ranged from .77 to .89.

Cognitive ability

We retrieved cognitive ability scores from archival records of the Flemish medical admission exam. The Gf test used in this exam consisted of 50 items (verbal, numeric, or figural) with five possible answers designed to tap general inductive and deductive reasoning. Internal consistency reliability was .72. As these scores were gathered during the admission exam, they were range-restricted. Range restriction refers to the phenomenon that when selecting people from a population (e.g., as part of an admission exam) the variability in their test scores is reduced as only the admitted students pursue education, thereby artificially reducing the magnitude of the correlation coefficients. Hence, we used the multivariate range restriction formulas of Ree, Carretta, Earles, and Albert (1994) to appropriately correct our correlation matrix, which then served as input for all analyses.

ANALYSES AND RESULTS

Table 1 presents the means, standard deviations, and range restriction corrected correlations among

the study variables. According to Cohen's (1988) criteria, correlations between .10 and .30, .30 and .50, and .50 and .70 can respectively be defined as small, moderate, and large. H1 posited small correlations between the STEU and the STEM and the personality traits. Supporting H1, the average correlations of $r = .04$ and $r = .07$ for the STEU and STEM respectively are indicative of independence from personality. Only openness and agreeableness showed somewhat stronger associations with the STEU and the STEM.

H2 posited the STEU and the STEM to show moderate correlations with Gf. However, the STEU and the STEM showed respectively a small and no significant correlation with Gf. The correlations between the EI tests and personality are also not significantly different from the correlations between the EI tests and Gf: $t(761) = -1.13$, $p = .260$ for the STEU and $t(761) = 0.34$, $p = .734$ for the STEM. These results are in line with MacCann and Roberts' (2008) and Austin's (2010) findings (see Appendix).

H3 stated that the STEU and the STEM would correlate more strongly with a performance-based EI test than with a self-report EI measure. Table 1 shows that the STEU and the STEM were more strongly and positively related to the performance-based EI measure than to the self-report EI measure, $t(761) = 3.58$, $p < .001$ for the STEU and $t(761) = 2.03$, $p < .05$ for the STEM.

To further test our hypotheses, we used CFA procedures via EQS (Bentler, 1995) through maximum-likelihood estimation. 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), the Bollen's incremental fit index (IFI) and the comparative fit index (CFI) to have values $> .95$, and the root mean square error of approximation (RMSEA) to be $< .05$. Finally, we also used Akaike's information criterion (AIC), with smaller values representing a better fit (Hu & Bentler, 1995).

For scrutinizing the relationship of the STEU and the STEM with personality and Gf, two competing measurement models were tested, hypothesizing the EI SJTs to load on the same factor as (1) Gf and (2) personality. To conceptualize personality in our models, we used one personality trait indicated by its three parcels and repeated all analyses with the other personality traits, because the model wherein all five personality trait scores were posited to load on one factor produced an unacceptable fit to the data. So, we tested five EI-as-intelligence models considering the EI SJTs to reflect mental performance rather

TABLE 1
Means (M), standard deviations (SD), and range restriction corrected correlations among study variables (N = 764)

	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Fluid cognitive ability	6.05	1.14	(.72)												
2. Extraversion	3.38	0.63	-.10*	(.88)											
3. Agreeableness	4.11	0.39	-.12**	.18**	(.78)										
4. Emotional stability	3.22	0.70	.14**	.10*	.02	(.89)									
5. Openness	3.56	0.47	.16**	.23**	.06	.11*	(.77)								
6. Conscientiousness	3.46	0.64	-.06	.00	.23**	-.04	.02	(.84)							
7. Self-perceiving (SREIS)	14.82	1.91	-.11*	.20**	.38**	.01	.15**	.20**	(.67)						
8. Self-using (SREIS)	10.78	2.21	-.11*	.14**	.37**	-.20**	-.10*	-.02	.20**	(.84)					
9. Self-understanding (SREIS)	12.32	2.74	-.03	.29**	.21**	-.04	.47**	.10*	.30**	.15**	(.84)				
10. Self-managing (SREIS)	14.84	1.86	-.07	.36**	.44**	.09*	.20**	.09*	.39**	.16**	.29**	(.72)			
11. STEU	27.67	3.40	.10*	.02	.06	.00	.12**	-.01	.00	.09*	.08*	.03	(.72)		
12. STEM	141.17	4.62	.05	.00	.20**	.02	.05	.08*	.07	.10*	.08*	.06	.24**	(.85)	
13. MSFDE	32.87	4.20	-.02	.02	.13**	-.04	.03	.03	.02	.09*	.04	.06	.22**	.18**	(.68)

Reliabilities are reported in parentheses on the diagonal. Listwise deletion of missing data. * $p < .05$; ** $p < .001$.

TABLE 2
Summary of fit statistics of intelligence and personality measurement models tested

<i>Measurement models</i>	χ^2	<i>df</i>	χ^2/df	<i>AIC</i>	<i>IFI</i>	<i>CFI</i>	<i>RMSEA</i>
<i>EI-as-intelligence model</i>							
1: Factor 1: Personality—Extraversion	71.760	19	3.777	33.760	.971	.971	.060
Factor 2: STEU, STEM, verbal, numeric, and figurative intelligence							
2: Factor 1: Personality—Openness	107.349	19	5.650	69.349	.912	.911	.078
Factor 2: STEU, STEM, verbal, numeric, and figurative intelligence							
3: Factor 1: Personality—Agreeableness	105.523	19	5.554	67.523	.916	.916	.077
Factor 2: STEU, STEM, verbal, numeric, and figurative intelligence							
4: Factor 1: Personality—Conscientiousness	78.119	19	4.112	40.119	.957	.957	.064
Factor 2: STEU, STEM, verbal, numeric, and figurative intelligence							
5: Factor 1: Personality—Emotional stability	64.322	19	3.385	26.322	.975	.975	.056
Factor 2: STEU, STEM, verbal, numeric, and figurative intelligence							
<i>EI-as-personality model</i>							
1: Factor 1: STEU, STEM, Personality—Extraversion	77.650	19	4.087	39.650	.968	.968	.064
Factor 2: Verbal, numeric, and figurative intelligence							
2: Factor 1: STEU, STEM, Personality—Openness	103.583	19	5.452	65.583	.916	.915	.076
Factor 2: Verbal, numeric, and figurative intelligence							
3: Factor 1: STEU, STEM, Personality—Agreeableness	77.423	19	4.075	39.423	.944	.943	.064
Factor 2: Verbal, numeric, and figurative intelligence							
4: Factor 1: STEU, STEM, Personality—Conscientiousness	79.871	19	4.204	41.871	.956	.956	.065
Factor 2: Verbal, numeric, and figurative intelligence							
5: Factor 1: STEU, STEM, Personality—Emotional stability	70.633	19	3.718	32.633	.972	.972	.060
Factor 2: Verbal, numeric, and figurative intelligence							

AIC = Akaike information criterion; IFI = Bollen's incremental fit index; CFI = comparative fit index; RMSEA = root mean square error of approximation.

TABLE 3
Summary of fit statistics of emotional intelligence (EI) measurement models tested

<i>Measurement models</i>	χ^2	<i>df</i>	χ^2/df	<i>AIC</i>	<i>IFI</i>	<i>CFI</i>	<i>RMSEA</i>
<i>EI-as-performance based EI measure model</i>							
(Factor 1: Self-perceiving, self-using, self-understanding, and self-managing)	16.591	13	1.276	-9.409	.991	.990	.019
(Factor 2: STEU, STEM, and MSFDE)							
<i>EI-as-self report EI measure model</i>							
(Factor 1: STEU, STEM, self-perceiving, self-using, self-understanding, and self-managing)	105.723	14	7.552	77.723	.757	.752	.093
(Factor 2: MSFDE)							

AIC = Akaike information criterion; IFI = Bollen's incremental fit index; CFI = comparative fit index; RMSEA = root mean square error of approximation.

than preferred ways of behaving, and five EI-as-personality models to reflect preferred ways of behaving rather than mental performance. To support that the EI SJTs show small correlations with personality (H1) and moderate correlations with Gf (H2), the EI-as-Intelligence model should show a better fit to the data than the EI-as-personality model. Table 2 shows that the two models produced a similar fit to the data (factor loadings for the EI SJTs were below .23 for all models; for more information the first author can be contacted), refuting the possibility that the EI SJTs diverge more from personality than from Gf, which confirms our correlation results.

For scrutinizing the relationship of the STEU and the STEM with the self-report and the performance-based EI measures, two competing measurement models were tested. Table 3 shows that the model hypothesizing the EI SJTs to load on the same factor as the performance-based EI measure produced the best fit to the data, supporting H3. This model also produced the strongest factor loadings for the EI SJTs.

DISCUSSION

The current study sought to examine whether the STEU and the STEM could be considered EI

measures assessing emotional abilities. The established nomological network showed that the EI SJTs do not capture the same construct as personality and self-report EI measures, but correspond rather to performance-based EI measures. These results lend support to the feasibility of the SJT paradigm as a valid complementary approach for assessing emotional abilities. For example, when researchers or practitioners do not want to use self-report EI measures or when performance-based EI measures are too complex to administer, EI SJTs can provide a valuable complement to measure EI branches in a diversity of settings and for comparing the results to extant EI ability measures (e.g., the MSCEIT).

However, a caveat is in order. The correlations of the EI SJTs with the performance-based EI measure were small (.22 and .18). One possible explanation might be the use of one specific performance-based EI measure. We chose an emotion perception test because this is one of the strongest EI ability branches. Additionally, linking the EI SJTs with an emotion perception test constitutes a more stringent test of H3 than linking the EI SJTs with, for example, an emotion management test (e.g., Branch 4 of the MSCEIT) because such tests also typically include situational items. So, our results on the relationship between SJT measures and ability measures of EI are probably conservative. Hence, future studies with other EI ability measures are needed to confirm our results. As another possible explanation, there might have been restriction of range in the ability EI (similar to cognitive ability) among medical students, leading to the smaller correlations with the performance-based EI measures.

A second important finding from our study is that the relations between the SJTs and Gf were weak, indicating that the SJTs do not reflect the ability to think logically and solve problems in new situations, independent of acquired knowledge. This new finding brings up the question of whether the SJTs primarily assess emotional knowledge (i.e., Gc; see MacCann & Roberts, 2008) or alternatively represent a latent factor distinct from Gf and Gc (see MacCann, 2010).

This study has the following limitations. First, some generalizability concerns should be acknowledged. Our sample consisted of Flemish medical students. As we used listwise deletion, a group of participants (10%) were also removed from the main analyses, potentially leading to non-response bias. Future research needs to examine the construct-related validity of the STEU and the STEM in other samples and settings. Second, all our results are based on two SJT EI measures

that have a knowledge response instruction format, multiple-choice answers, and a specific scoring key. Research has found that these SJT features might impact on the results found (Lievens, Sackett, & Buyse, 2009; MacCann & Roberts, 2008; McDaniel et al., 2007).

We see the following fruitful avenues for future research. First, formats other than paper-and-pencil should be used to develop EI SJTs. When respondents are able to observe the verbal and nonverbal behavior of characters in the situation portrayed in a video-based SJT, the test becomes even more realistic, which might benefit its predictive potential. Also, video-based SJTs might be more appropriate to assess the fluid, experiential component of emotion (i.e., the ability to appropriately *respond* to emotional situations; Ortony, Revelle, & Zinbarg, 2007).

Second, future studies might try to disentangle methods and constructs (Arthur & Villado, 2008). One possibility consists of holding the construct (an EI branch) constant and varying the method (different EI measurement approaches). For example, the ability to understand and manage emotions might be measured via the SREIS (Brackett et al., 2006), the STEU and STEM respectively (MacCann & Roberts, 2008), and the MSCEIT (Mayer et al., 2002). Such research designs focus on convergent validity and enable one to answer such key questions as: How well do these different methods converge in assessing EI? How much variance is accounted for by method factors and by substantive construct factors respectively?

In conclusion, our results provide some support for the theoretical rationale behind the development of EI SJTs, and elucidate the value of the STEU and the STEM as complementary measures of individuals' emotional abilities, thereby broadening the measurement paradigms available in EI research and practice.

Manuscript received September 2011
Revised manuscript accepted March 2012
First published online July 2012

REFERENCES

- Arthur, W., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology, 93*, 435–442.
- Austin, E. J. (2010). Measurement of ability emotional intelligence: Results for two new tests. *British Journal of Psychology, 101*, 563–578.
- 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, Canada.)
- Bentler, P. M. (1995). *EQS: Structural equations program manual*. Encino, CA: Multivariate Software.
- Blickle, G., Momm, T. S., Kramer, J., Mierke, J., Liu, Y. M., & Ferris, G. R. (2009). Construct and criterion-related validation of a measure of emotional reasoning skills: A two-study investigation. *International Journal of Selection and Assessment*, 17, 101–118.
- Brackett, M. A., Rivers, S. E., 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.
- Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action*. New York, NY: Elsevier.
- Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, 63, 83–117.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology*, 70, 662–680.
- Dunning, D., Heath, C., & Suls, J. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5, 69–106.
- 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.
- Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde et al. (Eds.), *Personality psychology in Europe* (pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.
- Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Thousand Oaks, CA: Sage.
- Joseph, D. L., & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal of Applied Psychology*, 95, 54–78.
- Lane, R. D., Quinlan, D. M., Schwartz, G. E., Walker, P. A., & Zeitlin, S. B. (1990). The levels of emotional awareness scale: A cognitive–developmental measure of emotion. *Journal of Personality Assessment*, 55, 124–134.
- 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.
- Lievens, F., Sackett, P. R., & Buyse, T. (2009). The effects of response instructions on situational judgment test performance and validity in a high-stakes context. *Journal of Applied Psychology*, 94, 1095–1101.
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling*, 9, 151–173.
- MacCann, C. (2010). Further examination of emotional intelligence as a standard intelligence: A latent variable analysis of fluid intelligence, crystallized intelligence, and emotional intelligence. *Personality and Individual Differences*, 49, 490–496.
- MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion*, 8, 540–551.
- Mayer, J., Caruso, D., & Salovey, P. (1999). Emotional intelligence meets traditional standards for an intelligence. *Intelligence*, 27, 267–298.
- Mayer, J., Roberts, R. D., & Barsade, S. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology*, 59, 507–536.
- Mayer, J., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey, & D. J. Sluyter (Eds.), *Emotional development and emotional intelligence: Educational implications* (pp. 3–34). New York, NY: Basic Books.
- Mayer, J., Salovey, P., & Caruso, D. (2002). *The Mayer, Salovey, and Caruso Emotional Intelligence Test: Technical manual*. Toronto, Canada: Multi-Health Systems.
- McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, 60, 63–91.
- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, 75, 640–647.
- Ortony, A., Revelle, W., & Zinbarg, R. (2007). Why emotional intelligence needs a fluid component. In G. Matthews, M. Zeidner, & R. D. Roberts (Eds.), *The science of emotional intelligence: Knowns and unknowns* (pp. 288–304). New York, NY: Oxford University Press.
- Petrides, K. V. (2010). Trait emotional intelligence theory. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 3, 136–139.
- Ree, M. J., Carretta, T. R., Earles, J. A., & Albert, W. (1994). Sign changes when correcting for range restriction: A note on Pearson's and Lawley's selection formulas. *Journal of Applied Psychology*, 79, 298–301.
- Roseman, I. J. (2001). A model of appraisal in the emotion system: Integrating theory, research, and applications. In K. R. Scherer, & A. Schorr (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 68–91). New York, NY: Oxford University Press.
- Salovey, P., & Mayer, J. (1990). Emotional intelligence. *Imagination, Cognition, and Personality*, 9, 185–211.
- Schulze, R., Wilhelm, O., & Kyllonen, P. C. (2007). Approaches to the assessment of emotional intelligence. In G. Matthews, M. Zeidner, & R. D. Roberts (Eds.), *The science of emotional intelligence: Knowns and unknowns* (pp. 199–229). New York, NY: Oxford University Press.
- Wagner, S. M., & Kemmerling, R. (2010). Handling nonresponse in logistics research. *Journal of Business Logistics*, 31, 357–381.

Weekley, J. A., & Ployhart, R. E. (2006). *Situational judgment tests: Theory, measurement and application*. Mahwah, NJ: Lawrence Erlbaum Associates.

Weekley, J. A., Ployhart, R. E., & Holtz, B. C. (2006). On the development of situational judgment tests: Issues in item development, scaling, and scoring. In J. Weekley, & R. E. Ployhart (Eds.), *Situational judgment tests: Theory, measurement and application* (pp. 157–182). Mahwah, NJ: Lawrence Erlbaum Associates.

Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review, 19*, 188–202.

Wilhelm, O. (2005). Measures of emotional intelligence: Practice and standards. In R. Schulze, & R. D. Roberts (Eds.), *Emotional intelligence: An international handbook* (pp. 131–154). Cambridge, MA: Hogrefe & Huber.

APPENDIX

Comparison of nomological network established in Belgium (current study) to nomological networks established in Australia (MacCann & Roberts, 2008) and in the UK (Austin, 2010)

	<i>Current study</i>		<i>Comparison data^a</i>			
	<i>STEU</i>	<i>STEM</i>	<i>STEU</i>		<i>STEM</i>	
Crystallized intelligence	–	–	.49**	.32***	.41**	.10
Fluid cognitive ability ^b	.10*	.05	–	.12	–	.07
Extraversion	.02	.00	–.06	–	.09	–
Agreeableness	.06	.20***	.16*	–	.24*	–
Emotional stability	.00	.02	.04	–	–.02	–
Openness	.12***	.05	.11	–	–.05	–
Conscientiousness	–.01	.08*	–.02	–	–.05	–
Self-report EI measure ^c	.05	.08*	–	–.04 (.03)	–	.13 (.12)
Emotion perception task ^d	.22***	.18***	–	.30*** (.21**)	–	.07 (.13)

* $p < .05$; ** $p < .01$; *** $p < .001$. ^aComparison data from MacCann and Roberts (2008) and Austin (2010) (Austin data in italics). ^bComparison data from Austin (2010) with a letter series test as fluid cognitive ability measure. ^cComparison data from Austin (2010) with a modified version of the Schutte EI scale and the TEIQue as self-report EI measures (TEIQue data in parentheses). ^dComparison data from Austin (2010) with a face blends and a sad faces IT task (see Austin, 2010, for a detailed description) as emotion perception measures (sad IT data in parentheses).