Validity and Reliability of a Brief Emotional Intelligence Scale (BEIS-10)

Kevin A. Davies, Andrew M. Lane, Tracey J. Devonport, and Jamie A. Scott

University of Wolverhampton, Walsall, UK

Abstract. This study describes the development and validation of a brief self-report measure of emotional intelligence based on Salovey and Mayer’s (1990) conceptualization. In stage one, the 33-item Emotional Intelligence Scale (EIS: Schutte et al., 1998) was assessed for content validity by a panel of experts. The panel deemed 17 items unsuitable for further analysis. In stage two, a theoretically derived 5-factor solution and a unidimensional model were subjected to confirmatory factor analysis (CFA) in a student-athlete sample (n = 955). Results supported the multidimensional solution. The Brief Emotional Intelligence Scale (BEIS-10) was developed by extracting the two items from each factor with the most salient factor loadings. CFA results yielded good fit indices for the 10-item, 5-factor solution. Finally, stage three provided evidence of test-retest stability for the BEIS-10 over a 2-week period in a sample of 111 student-athletes. The BEIS-10 is offered as a valid and reliable measurement tool that has particular utility in situations where brevity is important.

Keywords: trait emotional intelligence, measurement, self-report, confirmatory factor analysis, brief measure

Emotional intelligence (EI), a term used to describe adaptive interpersonal and intrapersonal emotional functioning (Kirk, Schutte, & Hine, 2008) has emerged as a potentially important construct in psychological research (Matthews, Zeidner, & Roberts, 2004). The intuitive appeal of EI relates to the notion that people differ in measurable ways in terms of the emotional skills they possess, and that such individual differences may account for the variance on real world criteria (Austin, Saklofske, & Egan, 2005). Meta analysis studies have thus far provided good supporting evidence for the predictive utility of EI against a range of health-related (e.g., physical and mental health) (Schutte, Malouff, Thorsteinsson, Bhullar, & Rook, 2007) and performance-related variables (e.g., academic achievement and occupational performance) (Van Rooy & Viswesvaran, 2004). However, such findings have been largely overshadowed by concerns regarding the validity and reliability of measures used to assess the construct (Brackett & Geher, 2006; Dawda & Hart, 2000; MacCann, Matthews, Zeidner, & Roberts, 2003; Sjöberg, 2001).

A pertinent issue impeding the development of valid EI measures concerns the lack of consensus regarding the EI construct (Brackett & Geher, 2006). EI has been conceptualized both in terms of a mixed model that refers to a cluster of cognitive and noncognitive constructs relating to emotional functioning (Bar-On, 2001; Goleman, 1995), and as an ability model that refers to EI as a form of intelligence involving the processing of emotion (Mayer & Salovey, 1997; Salovey & Mayer, 1990). While considerable debate exists over the most appropriate operationalization of the construct, it is worthwhile noting that common features do exist. For example, all models emphasize elements involved in the recognition and regulation of one’s own emotions and the emotions of others, as well as the assimilation of emotions into thought (Matthews et al., 2004). These cognitive features could therefore be regarded as integral facets of the EI construct (Brackett & Geher, 2006).

With this in mind, it could be argued that Salovey and Mayer’s (1990) original ability-based model perhaps represents the most cohesive and comprehensive model of EI that lends itself to assessment (Schutte et al., 1998). Salovey and Mayer’s model postulates that EI consists of the following three categories of adaptive cognitive abilities: appraisal and expression of emotion, regulation of emotion, and utilization of emotions in problem solving. The first category consists of the components of appraisal and expression of emotion in the self and appraisal of emotion in others. The second category, regulation of emotions, contains the components of regulation of emotions in the self and utilization of emotions in others. The third category, utilization of emotion, includes the components of flexible planning, creative thinking, redirected attention and motivation. Mayer and Salovey (1997) also offered a revised process-oriented model of EI that emphasizes stages of development in EI, potential for growth, and the contributions emotions make to intellectual growth. However, it should be noted that the original model lends itself better to conceptualizing the various dimensions of an individual’s current state of emotional development. A scale based on a person’s current state of EI development arguably has
greater utility in theoretical research exploring the nature of EI, the effect of EI, and whether EI could be enhanced (Schutte et al., 1998).

Another pertinent issue regarding EI measurement concerns the method of assessment. Salovey and Mayer’s (1990) model conveys EI as a cognitive ability, thereby suggesting that the concept should be assessed using a criterion-based maximal performance measure (e.g., Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT); Mayer, Caruso, & Salovey, 1999). However, the operationalization of EI as an ability is problematic because of the subjectivity of emotional experience (Robinson & Clare, 2002). The main issue relates to the difficulty in creating items or tasks that can be scored accurately in a standard and objective manner. This is because the variable being measured is inherently subjective and involves the interpretation of emotions, which can be influenced by various factors such as cultural context, personal experience, and individual differences in emotional processing.

An alternative approach to the assessment of EI has been to conceptualize EI as a trait, measurable via self-report (Brackett & Geher, 2006; Schutte et al., 1998). Self-report measures of EI differ from performance-based measures in that they capture self-perceived emotional skills and abilities that are representative of typical performance opposed to maximal performance (Brackett & Geher, 2006). Indeed, this distinction may account for the modest correlations observed between self-report and performance-based measures of EI (Brackett & Mayer, 2003; Brackett & Salovey, 2006; Lopes, Salovey, & Straus, 2003).

Unlike performance-based measures, self-report measures are proposed to hold value in that they do not contradict the subjective nature of emotions (Petrides et al., 2006). However, self-report measures have received criticism due to their reliance on self-perceptions (Matthews et al., 2004). Self-perceptions may not be particularly accurate or even available to conscious interpretation making them vulnerable to the entire gamut of response sets and social desirability factors affecting self-report measures (e.g., deception and impression management) (Roberts et al., 2001). Nevertheless, it should be noted that such limitations are common to all scales based on self-report, including personality assessment, and therefore should not prohibit the utility of self-report EI measures. Moreover, collecting data relating to self-perceptions may be useful in itself for two reasons. First, if a person’s self-concept is accurate, perceived emotional skills and abilities may be predictive of actual skills and abilities (Brackett & Geher, 2006). Second, because people tend to act in accord with their stated beliefs (Bandura, 1997), self-reported scores may be predictive of actual performance in a particular context.

There exist many self-report measures of EI in the scientific and commercial literatures, but only a handful of measures have been developed with a clear theoretical foundation in mind. Consequently, many measures suffer serious conceptual and psychometric limitations (Pérez, Petrides, & Furnham, 2005). Among the most common weaknesses of extant measures is their incomplete coverage of the sampling domain of the construct, unstable factor structures, and salient loadings on the well-established personality factors (Davies, Stankov, & Roberts, 1998; Petrides & Furnham, 2001). In order to effectively use the theory of EI in research or to enhance practical applications, it is necessary to have an instrument that will accurately and efficiently assess the intended construct (Tapia & Marsh, 2006). Schutz (1994) argued that demonstrating existing measures are valid and reliable should be the first stage in the research process. Therefore, the present study sought to investigate the validity and reliability of one of the most widely used self-report measure of EI based on the Salovey and Mayer (1990) model: the Emotional Intelligence Scale (EIS) also known as the Schutte Self-Report Inventory (SSRI) (Schutte et al., 1998).

Emotional Intelligence Scale (Schutte et al., 1998)

The EIS is a 33-item self-report instrument purported to measure an individual’s perceptions of the extent to which s/he can appraise and regulate emotions in self and others, and utilize emotions for problem solving (Schutte et al., 1998). In the initial validation study, Schutte et al. used a set of 62 items derived from the model of Salovey and Mayer (1990). Exploratory factor analysis on data from 346 participants yielded a 4-factor model. The authors argued that by removing 29 items and reanalyzing the data, an adequate 1-factor solution was produced. Schutte et al. reported adequate internal consistency reliability (r = .87 to .90) and acceptable test-retest reliability (r = .78) for the unidimensional scale. Furthermore, there was evidence that the EIS was distinct from established measures of the Big Five.

Subsequent validation studies rightly questioned the unidimensional structure of the EIS and sought to establish a multifactorial solution. For instance, Petrides and Furnham (2000) sought to test the hypothesis, via confirmatory factor analysis (CFA), that the EIS measured a general factor of EI. Results indicated that the general factor model provided a poor fit to the data (e.g., comparative fit index: CFI = .51, root mean square error of approximation: RMSEA = .11). Consequently, Petrides and Furnham (2000) followed their CFA analysis with an unrestricted (exploratory) principal components analysis and extracted four components: optimism/mood regulation, appraisal of emotions, social skills, and utilization of emotions. Saklofske, Austin, and Minski (2003) subjected the EIS to CFA and found moderate support for the 4-factor model.

Gignac, Palmer, Manocha, and Stough (2005) recog-
nized the limitations associated with the use of exploratory data-driven procedures to demonstrate factorial validity (see Thompson & Daniel, 1996) and sought to establish whether a theoretically derived solution would yield improved psychometric properties. Following an assessment of content validity, 28 items representative of six dimensions of EI were extracted for further analysis. CFA results suggested that the hypothesized 6-factor model could not be completely recovered. Specifically, a nested-factors model with a first-order general factor, and four nested factors corresponding to appraisal of emotions in the self, appraisal of emotions in others, emotional regulation of the self, and utilization of emotions in problem solving were identified, in conjunction with a first-order acquiescence factor. There was no evidence to suggest the existence of an independent emotional expression or emotional regulation of others factor. Gignac et al. (2005) argued that further validation work on the scale was needed if the intention of the scale was to assess the theoretical model proposed by Salovey and Mayer (1990).

Recently, Lane et al. (2009) sought to expand on the work of Gignac et al. (2005) by establishing a factor structure consistent with theoretical proposals of Salovey and Mayer (1990). A principle reason for investigating the EIS offered by Lane et al. was to determine the validity of the scale for use with an athletic population. Lane et al. scrutinized the 33- items of the EIS for content validity and categorized them into one of six dimensions based on EI theory (appraisal of own emotions, appraisal of others’ emotions, regulation of own emotions, social skills/regulation of others’ emotions, utilization of emotion, and optimism). CFA yielded a poor fit to the data for a single-factor solution typically used in the literature and an acceptable fit to the data for the theoretically derived 6-factor solution. In an attempt to improve model fit, Lane and colleagues respecified the model: Using CFA results in conjunction with content validity results they removed 13 items that contained no emotional content. For example, “I motivate myself by imagining a good outcome to tasks I take on” and “I know when to speak about my personal problems to others.” Lane et al. (2009) argued that if definition items that assess EI should contain reference to feelings in general and references to specific emotions. As the above procedure removed all but 1 item from the optimism scale, this factor was discarded from further analyses. Subsequent CFA on the remaining 19 items resulted in acceptable fit indices for the single-factor solution and good fit indices for the 5-factor solution.

The multifactorial solution offered by Lane et al. (2009) holds promise as it closely resembles Salovey and Mayer’s (1990) conceptualization of EI. Such a factor structure permits the examination of hypothesis that specify relationships between facets of EI and other outcome variables. One potential limitation of Lane et al.’s model however, concerns the uneven number of items per factor. For example, while utilization of emotion is represented by 6 items, regulation of own emotions is represented by just 2 items. Anastasi and Urbina (1997) argue that the items in an assessment instrument should be distributed, or weighted, in a way that reflects the relative importance of the various facets of the targeted construct. If items overrepresent or underrepresent facets of a construct, the obtained scores and inferences from these scores will be biased. Given that facets of EI are deemed equally important, we suggest that Lane et al.’s model can be improved by producing factors with an equal number of indicators.

Length of Construct

An important decision in the development of a questionnaire concerns the appropriate number of items to include in each factor. There are two arguments to consider. First, from a theoretical perspective, it has been suggested that there are potentially an infinite number of indicators to represent any given concept (Anastasi & Urbina, 1997). Logically, each item represents a part of the construct, and therefore the more items assessed, the greater the chance the construct will be accurately assessed. However, in practical terms, a large number of items can sometimes be counterproductive, and many concepts may be understood using just one or two items. For example, in developing a brief measure of the Big Five Personality Test, Gosling, Rentfrow, and Swann (2003) proposed that, when a concept such as extroversion is widely understood, it is more straightforward to ask the person how extroverted they are rather than asking them how talkative, outgoing, gregarious, and enthusiastic they are. Gosling et al. (2003) developed a 5-item and 10-item measure of the Big Five. Although both measures reported adequate validity and reliability, the 10-item measure demonstrated superior psychometric properties.

The second argument considers statistical methods for the assessment of validity. Watson and Clark (1997) argued that factors with less than 4 items typically fail to yield an internal consistency (alpha) coefficient above the generally accepted criterion value of 0.70. However, Schutz and Gesseroli (1993) and Lane (2007) are highly critical of the use of alpha values to support factorial validity and suggest that more important information derives from CFA data. Based on the lower limit of identifiability, Kline (2005) suggests that CFA requires a minimum of 3 items per factor. Bollen (1989), however, contends that as long as there are multiple factors and the factors are not independent, solutions with two indicators per factor are best. Marsh, Hau, Balla, and Grayson (1998) recommend that when there are only 2 items per factor, large sample sizes should be used (> n = 400) to be confident about obtaining a fully proper solution. Considering both arguments, if concepts as broad and complex as those identified in the Big Five of personality can be adequately represented by just 2 items, then it is argued that conceptually smaller concepts such as EI could follow this trend.
Utility of a Brief Self-Report Measure of EI

Research has typically employed measures of EI in large-scale correlational studies intended to address multiple hypotheses within the sample (Zeidner, Roberts, & Matthews, 2008). While this approach has been used to maximize the use of outcome data and makes optimal use of participants’ involvement, the drawback relates to the increased response burden placed on participants, that is, the amount of time and effort demanded by a project can be substantial. Similarly, internet-based surveys have become an increasingly popular method of data collection in psychology (Lonsdale, Hodge, & Rose, 2006), although it has been suggested that respondents are unlikely to tolerate the same degree of response burden as seen in face-to-face methods (Granello & Wheaton, 2004). Given that a large response burden is likely to lead to lower levels of accuracy, commitment and adherence to studies, it is proposed that a brief measure of EI may broaden the scope for future research.

Purpose of Research

In summary, research suggests that the content validity and factorial validity of the 33-item EIS is questionable. Given that Lane et al. (2009) retained a factor structure congruent with Salovey and Mayer’s (1990) conceptualization of EI by removing theoretically redundant items, it is proposed that a brief version of the EIS could reasonably serve as a more efficient measure of EI. Accordingly, this article presents preliminary validity and reliability data for a brief version of the EIS. The research was conducted in three stages.

Stage 1: Content Validity

Content validity refers to the extent to which items represent the construct they are purported to measure. A standard approach to establishing content validity is to use experts to select or confirm items that best describe the construct in question (Lane et al., 2009; Worthington & Whittaker, 2006).

Method

Participants and Procedure

The first three authors, who had all published research on EI in peer-refereed academic journals, scrutinized each of the 33 EIS items independently, assessing them first for affective content (the item had to contain a reference to emotion, feelings, or mood), and second for theoretical relevance in relation to Salovey and Mayer’s (1990) model. Items that satisfied both criteria were retained for empirical analysis.

Stage 2: Factorial Validity

Factorial validity is the degree to which a measure of a construct conforms to the theoretical definition of the construct and is considered an important component of establishing evidence for the validity of inferences from test scores. Factorial validity is established by testing the fit of a theoretically based measurement model for describing the variances and covariances underlying items on a scale using CFA (Tabachnick & Fidell, 2007).

Table 1. Content validity results for the 33 items from the EIS

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appraisal of own emotions</td>
<td>9, 19, 22</td>
</tr>
<tr>
<td>Appraisal of others’ emotions</td>
<td>18, 26*, 29, 32, 33</td>
</tr>
<tr>
<td>Regulation of own emotions</td>
<td>11*, 12*, 14, 21, 26*</td>
</tr>
<tr>
<td>Regulation of others’ emotions</td>
<td>11*, 13, 30</td>
</tr>
<tr>
<td>Utilization of emotion</td>
<td>7, 12*, 17, 20, 27, 31</td>
</tr>
<tr>
<td>Expression of emotion</td>
<td>11*</td>
</tr>
<tr>
<td>No emotional content</td>
<td>1, 2, 3, 4, 5, 6, 10, 15, 16, 23, 24, 25, 28</td>
</tr>
<tr>
<td>Unclassified</td>
<td>8</td>
</tr>
</tbody>
</table>

*Item assesses more than one dimension of EI.
Method

Participants

A total of 955 student-athletes volunteered to participate in the study; thereof 496 specified their sex as male (aged 21.2 ± 3.7 years) and 395 as female (aged 21.1 ± 4.4 years). The participants were drawn from three higher education universities across the UK.

Measurement Instrument

The 16 items (retained from the EIS) are rated on a 5-point likert scale anchored by 1 = strongly agree to 5 = strongly disagree. One item is reverse scored.

Procedure

Following institutional ethical approval, we obtained informed consent from all participants. All agreed to take part in the study voluntarily, meaning no incentives were offered. The retained items were administered as part of a larger survey exploring coping among student-athletes. Participants completed either a paper-based version of the survey during a timetabled lecture, or a web-based version at their own leisure. Instructions to participants included a reminder to answer all questions honestly and independently, and a statement to reduce social desirability bias.

Data Analysis

Confirmatory factor analyses were performed using EQS V6.1 (Bentler & Wu, 2003). Two models were tested using maximum likelihood estimation. The first was a 16-item, 5-factor solution that sought to distinguish appraisal of own emotions, appraisal of others’ emotions, regulation of own emotions, regulation of others’ emotions, and utilization of emotions. The model specified that items were related to their hypothesized factor with the variance of the factor fixed at 1. Factors were allowed to freely intercorrelate. The second model tested was a 16-item, single-factor first-order model. This model was tested as research has typically summed EIS scores to produce a single score. Model fit was assessed using two incremental fit indices, the comparative fit index (CFI: Bentler, 1990) and the non-normed fit index (NNFI) or Tucker-Lewis index (TLI: Tucker & Lewis, 1973). Incremental fit indices are based on comparisons between the hypothesized model and a null model (in

<table>
<thead>
<tr>
<th>No</th>
<th>Items</th>
<th>Five-factor model</th>
<th>One-factor model</th>
<th>BEIS-10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Factor loading</td>
<td>Error variance</td>
<td>Factor loading</td>
</tr>
<tr>
<td>9</td>
<td>I am aware of my emotions as I experience them*</td>
<td>.49</td>
<td>.87</td>
<td>.47</td>
</tr>
<tr>
<td>19</td>
<td>I know why my emotions change</td>
<td>.63</td>
<td>.78</td>
<td>.58</td>
</tr>
<tr>
<td>22</td>
<td>I easily recognize my emotions as I experience them</td>
<td>.74</td>
<td>.67</td>
<td>.62</td>
</tr>
<tr>
<td>32</td>
<td>I can tell how people are feeling by listening to the tone of their voice</td>
<td>.71</td>
<td>.70</td>
<td>.60</td>
</tr>
<tr>
<td>18</td>
<td>By looking at their facial expressions, I recognize the emotions people are experiencing</td>
<td>.72</td>
<td>.70</td>
<td>.62</td>
</tr>
<tr>
<td>29</td>
<td>I know what other people are feeling just by looking at them*</td>
<td>.70</td>
<td>.71</td>
<td>.62</td>
</tr>
<tr>
<td>33</td>
<td>It is difficult for me to understand why people feel the way they do*</td>
<td>.18</td>
<td>.98</td>
<td>.17</td>
</tr>
<tr>
<td>14</td>
<td>I seek out activities that make me happy</td>
<td>.70</td>
<td>.71</td>
<td>.54</td>
</tr>
<tr>
<td>21</td>
<td>I have control over my emotions</td>
<td>.53</td>
<td>.85</td>
<td>.42</td>
</tr>
<tr>
<td>13</td>
<td>I arrange events others enjoy</td>
<td>.56</td>
<td>.83</td>
<td>.45</td>
</tr>
<tr>
<td>30</td>
<td>I help other people feel better when they are down</td>
<td>.74</td>
<td>.67</td>
<td>.58</td>
</tr>
<tr>
<td>7</td>
<td>When my mood changes, I see new possibilities*</td>
<td>.37</td>
<td>.93</td>
<td>.32</td>
</tr>
<tr>
<td>17</td>
<td>When I am in a positive mood, solving problems is easy for me*</td>
<td>.58</td>
<td>.82</td>
<td>.52</td>
</tr>
<tr>
<td>20</td>
<td>When I am in a positive mood, I am able to come up with new ideas</td>
<td>.59</td>
<td>.81</td>
<td>.54</td>
</tr>
<tr>
<td>27</td>
<td>When I feel a change in emotions, I tend to come up with new ideas*</td>
<td>.58</td>
<td>.82</td>
<td>.52</td>
</tr>
<tr>
<td>31</td>
<td>I use good moods to help myself keep trying in the face of obstacles</td>
<td>.63</td>
<td>.78</td>
<td>.58</td>
</tr>
</tbody>
</table>

*denotes the items removed from the 5-factor model due to weakest factor loading.
which there are no relationships among the observed variables) and are not influenced by sample size (Bentler, 1990). Although a value of > .90 was originally considered representative of a well-fitting model, a revised cutoff value close to .95 has been advised (Hu & Bentler, 1999). The third index used was the root mean square error of approximation (RMSEA; Steiger, 1990), which indicates the mean discrepancy between the observed covariances and those implied by the model per degree of freedom, and therefore has the advantage of being sensitive to model complexity. A value of .05 or lower indicates a good fit, and values up to .08 indicate an acceptable fit (Browne & Cudeck, 1993).

**Results**

CFA results for the 5-factor model were as follows: CFI = .91, NNFI = .89, and RMSEA = .06. Although both incremental fit indices were below the .95 criterion suggested by Hu and Bentler (1999) for a well-fitting model, a CFI greater than .90 may be considered acceptable (Bentler, 1992). The RMSEA was also an acceptable value. In contrast, fit indices for the single-factor model were all indicative of a poor-fitting model (CFI = .80, NNFI = .77, and RMSEA = .09). Factor loadings for the items on both models are contained in Table 2. In general, factor loadings support the hypothesized relationships between items in the 5-factor model, with just 2 items demonstrating a poor relationship (< .44) (See Comrey & Lee, 1992) with their hypothesized factor.

Next, in order to produce a more parsimonious measure (BEIS-10), the 5-factor model was respecified by retaining the 2 items with the most salient factor loading on their hypothesized factor. This decision ensured that all factors had an equal number of indicators. CFA results for the BEIS-10 suggested an improved fit to the data (CFI = .97, NNFI = .94, RMSEA = .06). An inspection of the standardized solution results (see Table 2) indicates that 3 items have factor loadings that can be considered excellent (> .71), 3 items very good (> .63), 3 items good (> .55), and 1-item fair (> .45) (Comrey & Lee, 1992).

Interfactor correlations (see Table 3) suggest that EI dimensions share a high degree of variance. However, based on the content validity data, we contend that the factors retain a degree of independence and therefore no higher-order model was tested.

<table>
<thead>
<tr>
<th>Table 3. Interfactor correlations for the BEIS-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>1 Appraisal of own emotions</td>
</tr>
<tr>
<td>2 Appraisal of others' emotions</td>
</tr>
<tr>
<td>3 Regulation of own emotions</td>
</tr>
<tr>
<td>4 Regulation of others' emotions</td>
</tr>
<tr>
<td>5 Utilization of emotion</td>
</tr>
</tbody>
</table>

*p < .05.

**Stage 3: Test-Retest Reliability of the BEIS-10**

Establishing reliability is vital to validating psychometric tools (Anastasi & Urbina, 1997). In order to be considered reliable, a measure must report scores on two occasions within a predetermined degree of accuracy (Kline, 2005). As the BEIS-10 is representative of an individual’s self-reported emotional skills and abilities, it is hypothesized that the measure should demonstrate a degree of stability over time (Petrides, Furnham, & Mavroveli, 2007).

**Method**

**Participants**

Participants were 111 undergraduate student-athletes from a UK University; thereof 61 specified their sex as male (aged 21.15 ± 4.0 years) and 48 as female (aged 21.7 ± 5.7 years). The sample size used in this study is commensurate with that recommended (N = 100) for assessing the reliability of a psychometric questionnaire (Kline, 2005; Nevill, Lane, Kilgour, Bowes, & Whyte, 2001).

**Procedure**

The BEIS-10 was administered under standardized conditions (at the start of a lecture) on two separate occasions separated by a 2-week time gap.

**Data Analysis**

Reliability was assessed using two methods. First, because scores between test-retest measurements on a psychometric questionnaire are discrete and hence not normally distributed, Nevill et al. (2001) recommend adopting a nonparametric approach to assessing agreement. Nevill and colleagues suggest calculating the proportion of differences within a reference value of ± 1 for each item. In order to be considered reliable, 90% of values should be between these reference points. The statistical software package SPSS was used to construct frequency tables. Second, the association between factor test-retest scores was calculated using Pearson’s product-moment correlations, which is the standard approach.

**Results**

Test-retest reliability scores are included in Table 4. The proportion of agreement scores for items ranged from 89.2% to 96.4% within a ± 1 range. All except 1 item achieved the 90% criteria suggested by Nevill et al. (2001)
for assessing the reliability of a stable self-report measure. Correlations for factor test-retest scores indicate significant moderate positive correlations between all factor scores over a 2-week period.

**Discussion**

Previous research exploring the factorial validity of the EIS often failed to produce a factor structure congruent with theory (Petrides & Furnham, 2000; Saklofske et al., 2003; Schutte et al., 1998). This resulted in questions regarding the utility of the measure. Using theory-driven methods (see Schutz, 1994), we aimed to test the validity and reliability of a brief version of the EIS (BEIS-10) to establish whether it could serve as a more valid and efficient measure.

**Content Validity of the 33-Item EIS**

In stage one, items from the EIS were assessed by a panel of experts for content validity. Results identified several limitations with the 33-item measure. Most noticeably, 13 items lacked emotional content. By definition, trait EI refers to “emotion related self-perceptions and behavioral dispositions relating to the perception, processing, and utilization of emotion-laden information” (Petrides & Furnham, 2001, p. 426). Although an item such as “I find it hard to understand the nonverbal messages of other people” may appear to relate to the perception and processing of others’ emotions, the item does not indicate that the other person is expressing emotion in a nonverbal manner, consequently emotional processing may not take place. We therefore suggest that the utility of such an item is questionable in the assessment of trait EI.

Three further items were identified as lacking independence on a single dimension of EI. Previous attempts to classify the same items under a single dimension of EI also produced conflicting results (see Gignac et al., 2005; Lane et al., 2009). In order to be considered suitable for assessment, items must retain a degree of independence on a single factor (Worthington & Whittaker, 2006). This ensures that factors have relative independence and practical utility when assessing relationships between components of EI and related constructs. Crossloading items are likely to result in model specification problems – and more importantly may impede theoretical understanding. In addition, a further item was deemed theoretically redundant as it was not representative of a dimension of EI as specified by Salovey and Mayer (1990). The remaining 16 items were deemed representative of the 5 dimensions of Salovey and Mayer’s (1990) model. This finding lends some credibility to Schutte et al. (1998) suggestion that the items in the unidimensional scale adequately cover the sampling domain of Salovey and Mayer’s (1990) model.

It is important to note that, similar to Gignac et al. (2005), the factor expression of emotion could not be identified. While items 1, 15, and 16 were deemed representative of verbal and nonverbal expression, they contained no emotional content making them unsuitable for inclusion. Furthermore, although the item “I like to share my emotion with others” could arguably be deemed representative of this factor, expression of emotion is likely to be dependent on contextual factors (Barrett & Gross, 2001). Indeed, emotion theorists have increasingly argued that whether one expresses or suppresses emotional expression is not as im-

<table>
<thead>
<tr>
<th>Item</th>
<th>Time 1 mean ($)</th>
<th>Time 2 mean ($)</th>
<th>PA ± 1 (%)</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appraisal of own emotions</td>
<td>I know why my emotions change</td>
<td>3.52 (.77)</td>
<td>3.61 (.78)</td>
<td>102 (91.8)</td>
</tr>
<tr>
<td></td>
<td>I easily recognize my emotions as I experience them</td>
<td>3.67 (.62)</td>
<td>3.70 (.61)</td>
<td>105 (94.5)</td>
</tr>
<tr>
<td>Appraisal of others’ emotions</td>
<td>I can tell how people are feeling by listening to the tone of their voice</td>
<td>3.66 (.72)</td>
<td>3.69 (.72)</td>
<td>99 (89.2)</td>
</tr>
<tr>
<td></td>
<td>By looking at their facial expressions, I recognize the emotions people are experiencing</td>
<td>3.79 (.65)</td>
<td>3.86 (.65)</td>
<td>104 (93.7)</td>
</tr>
<tr>
<td>Regulation of own emotions</td>
<td>I seek out activities that make me happy</td>
<td>3.93 (.72)</td>
<td>3.95 (.64)</td>
<td>104 (93.7)</td>
</tr>
<tr>
<td></td>
<td>I have control over my emotions</td>
<td>3.41 (.83)</td>
<td>3.44 (.75)</td>
<td>105 (94.5)</td>
</tr>
<tr>
<td>Regulation of others’ emotions</td>
<td>I arrange events others enjoy</td>
<td>3.59 (.91)</td>
<td>3.57 (.84)</td>
<td>100 (90.1)</td>
</tr>
<tr>
<td></td>
<td>I help other people feel better when they are down</td>
<td>3.93 (.68)</td>
<td>3.98 (.65)</td>
<td>103 (92.8)</td>
</tr>
<tr>
<td>Utilization of emotions</td>
<td>When I am in a positive mood, I am able to come up with new ideas</td>
<td>3.66 (.65)</td>
<td>3.77 (.66)</td>
<td>106 (95.4)</td>
</tr>
<tr>
<td></td>
<td>I use good moods to help myself keep trying in the face of obstacles</td>
<td>3.68 (.68)</td>
<td>3.70 (.58)</td>
<td>107 (96.4)</td>
</tr>
</tbody>
</table>

*p < .05.
portant for adjustment as is the ability to flexibly express or suppress emotional expression as demanded by the situational context (Barrett & Gross, 2001; Bonanno, 2001). For example, an emotionally intelligent individual would not choose to express emotion in a situation where it may have a negative impact on themselves and/or others. This item is therefore deemed unsuitable for inclusion in a dispositional measure such as the BEIS-10.

The failure to recover an emotional expression factor represents a reduction in width and depth of Salovey and Mayer’s (1990) EI construct. The flexible expression of emotion serves multiple adaptive functions, including communicating and regulating internal states (Ekman & Davidson, 1993), and developing and maintaining social interactions (Ekman, 1993). In recognition of the importance of this dimension, we propose two potential solutions. First, researchers could utilize preexisting measures that tap facets of emotional expression (e.g., Emotional Regulation Questionnaire: Gross & John, 2003; or the Berkeley Expressivity Questionnaire: Gross & John, 1997). However, it is important to note that many existing measures fail to take into account contextual factors that influence expression. Researchers adopting such an approach should therefore consider appropriate methodologies for collecting data (e.g., over a variety of situational contexts). Second, researchers could seek to develop items that are suitable for inclusion in a dispositional measure such as the BEIS-10. An example item could be “I know when it is appropriate to express my negative emotions to others.”

Although it is a commonly held belief that content validity is a key part of the validation process, it typically takes second place to examining factorial validity (Lane et al., 2009). Findings from the present study lend support to the value of exploring content validity and scrutinizing the intended meaning of items closely. The identification of 17 items from the EIS that are unsuitable for the assessment of EI offers a meaningful contribution to the validation of the measure. This finding may explain why previous attempts to establish a factor structure commensurate with Salovey and Mayer’s (1990) model have been unsuccessful.

Factorial Validity of the BEIS-10

A central aspect of the nature of EI is that it is concerned with regulatory processes related to one’s own emotions and the emotions of others (Gignac et al., 2005). A composite measure of EI (e.g., Petrides & Furnham, 2000) cannot distinguish emotions related to self from emotions related to others. Furthermore, broad factors (e.g., appraisal of emotion and regulation of emotion), as seen with the hierarchical structure of Salovey and Mayer’s (1990) model, are unhelpful as they mask relationships between facets of EI and other variables. For example, the skills needed to regulate one’s own emotions differ considerably from those needed to regulate others’ emotions. Regulation of others’ emotions requires social-based skills whereas regulation of one’s own emotions requires psychological skills. The theoretically derived factor structure retrieved in the present study is advantageous as it permits the examination of hypotheses that are seeking to understand relationships between facets of EI and other outcome variables (Tapia & Marsh, 2006).

Reliability of the BEIS-10

The proportion of agreement results of the present study lend partial support to the relative stability of the BEIS-10 over a 2-week period. However, the relationships reported by way of Pearson’s correlation were more moderate than expected for a stable construct (< .8) (Kline, 2005). There are two potential explanations for this finding. First, it is possible that the measurement error associated with the instrument (e.g., unclear item content) led to a reduction in test-retest correlations because of random responses. However, given the thorough content validity process employed in the present study, we contend that such effects are likely to be minimal. Second, it is plausible that, while actual EI abilities are likely to remain relatively stable over time, similar to other forms of intelligence (e.g., IQ) self-perceptions of emotional skills and abilities may be more transient in nature. For example, individuals reporting depressed mood may be less confident in their ability to regulate others’ emotions than if they were experiencing a positive mood. With this in mind, rather than reflecting a weakness of the measure per se, this result may help explain day-to-day variations in interpersonal and intrapersonal functioning. Future research should seek to account for transient mood states when exploring the relationship between EI and measures of interpersonal and intrapersonal functioning.

Why Not Develop a New Measure?

Given the number of changes made to the scale, it could be argued that the research process should have started from theoretical principles rather than seeking to revise an existing scale. While such an approach is plausible, this was the approach adopted by EI researchers to date, resulting in a plethora of measures. Developing a new scale from principles can enhance the literature, but in developing the case for this approach researchers should first test the validity of an existing measure. As such, once a scale has been introduced into the literature, it is incumbent upon researchers to investigate the validity and reliability of the measure. Although such an approach might ostensibly seem unambitious in terms of knowledge enhancement, unless researchers conduct rigorous validation of their scales, knowledge is likely to go round in circles rather than forward. If researchers cannot trust the validity of their scales, findings from studies that use them cannot be relied upon.
Findings of the present study offer a meaningful contribution to the EI literature by identifying limitations in 17 EIS items and emphasizing the importance of assessing EI through a multifactor rather than a single-factor model.

Proposed Benefits of the BEIS-10

The central benefit of the BEIS-10 is that it offers a way to capture EI quickly while maintaining acceptable psychometric properties. At just 10 items, the measure is considerably shorter than any other measure of trait EI, taking on average between one to two minutes to complete. We suggest that the BEIS-10 may be particularly useful for collecting data in population groups and under conditions in which time is an issue. Indeed, the expansion in research in EI is in part driven by the notion that such individual differences may account for the variance on real world criteria (Austin et al., 2005). Brevity of measures has been highlighted as fundamental in research that seeks to maintain a high degree of ecological validity in settings such as academic examinations and athletic competition (Lane, 2007).

Limitations and Future Research

Content validity is conditional for a particular construct domain. Because there are several models of EI, it is important to recognize that the BEIS-10 was developed specifically with Salovey and Mayer’s (1990) conceptualization of EI in mind and as such may only have satisfactory content validity for this definition. Furthermore, although the theoretically based factor structure retrieved in the present investigation offers a number of potential advantages, we urge researchers from other domains to test the stability of the factor structure in other population groups.

It is also important to note the limitations associated with performing CFA on measures that contain just 2 items per factor. First, as noted by an anonymous reviewer, there are conflicting positions in the methodologists’ community as to whether only a complete structural equation model, or also the single, partial measurement models must be identified in order to allow interpretation of the analysis. Irrespective of viewpoint, we acknowledge that only the first condition can be satisfied with a measure that contains two indicators per factor. However, to be confident of obtaining a fully proper solution, we followed the recommendations of Marsh et al. (1998), who suggest using a large sample sizes (n ≥ 400). Second, fit is often exaggerated when factors contain just 2 items: Factor loadings are often too large, uniqueness too small, and factor correlations too large (Marsh et al., 1998). This may explain the high interfactor correlations observed in the present study. With these limitations in mind, it is important to emphasize the practical utility of the BEIS-10 opposed to the psychometric properties of the measure.

Because of the limited number of items in each factor, it is also important to recognize the limitations associated with the coverage of Salovey and Mayer’s (1990) EI construct. Very brief measures, such as the BEIS-10, do not provide scores for the narrower facet-level constructs. For example, utilization of emotions is comprised of flexible planning, creative thinking, redirected attention, and motivation. However, the two items in the BEIS-10 are only representative of creative thinking and motivation. This reduction in content may limit the scope of the BEIS-10 as it is more difficult to be certain of obtaining a “true score” (Marsh et al., 1998). Yet the primary purpose of the BEIS-10 is to collect EI data quickly. Practitioners seeking a more in-depth, accurate representation of facets of EI should consider alternative measures that use multiple indicators.

A recognized limitation of the current research program is that the criterion validity of the BEIS-10 was not explored. In order to have utility in the field of EI research, the BEIS-10 must demonstrate that it is independent from existing measures of personality and predictive of functional outcomes (e.g., well-being). Further research exploring this type of validity is clearly warranted.

Conclusion

The results of the present investigation suggest that the utility of the 33-item EIS is compromised by way of its poor content validity. However, a revised version of the measure, the BEIS-10, demonstrates evidence of content validity, factorial validity, and test-retest reliability. We propose that the BEIS-10 may adequately serve as a more efficient and psychometrically sound version of the EIS. The BEIS-10 is proposed to be particularly useful in situations where brevity is important and may help to maintain ecological validity.

References


Kevin A. Davies

University of Wolverhampton
Research Centre for Sport, Exercise, and Performance
Walsall Campus
Gorway Road
Walsall WS1 3BD
UK
Tel. +44 1902 323303
Fax +44 1902 323295
E-mail kad@wlv.ac.uk